Thesis for the degree of Canditatus Scientiarum

Gravitational Microlensing of Quasar Clouds

Detectability in a Worst-Case Scenario

Kjetil Kjernsmo

January 23, 2002

Institute of Theoretical Astrophysics
University of Oslo
Norway

Abstract

It is believed that the Broad Emission Line Region of Quasars consists of discrete clouds. Due to the suspected small size of a cloud, the radiation from clouds may be strongly amplified by gravitational microlensing.

It is shown that under some conditions, it is possible to detect gravitational microlensing even in a situation where the position and the velocity of clouds are non-correlated.

Recommendations are made with respect to planning of observations, based on whether one has knowledge as to the number of clouds or not. If one knows that there are approximately 1 million clouds, there is good chance to detect lensing. If one has no knowledge of the number of clouds, one should choose instruments and objects that can provide the strongest lines and resolve it into as many as 1000 bins.

Linear regression is used to derive relations that can be used for prediction of conditions where lensing can be detected.

Preface

When starting this study, I decided that I would define my own problem. This prompted a dive into research literature, and I came up with many interesting issues that I worked with for short periods, ranging from elaborate planning of optical monitoring of gravitational lens systems, via X-ray-studies, to quick-response programs for multi-wavelength monitoring of microlensing events.

I have no regrets about having looked at many different problems, however, as it gave me a better overview of the research literature.

When I finally decided that the subject for this thesis was going to be microlensing of the clouds of the Broad Emission Line Region, it was first the intention to aim at a physical understanding. As the work proceeded, however, the physics drifted further and further into the background, to be replaced by statistics. In the final result, the physics can be found mostly in the simplifying assumptions I make and in the discussion of microlensing-induced amplification.

This was hardly intentional, as my undergraduate studies were focused on physics and included very little training in statistics.

I have no regrets about this either, on the contrary, it prompted me to learn more about statistics, and I have (I hope) come to appreciate the background for the rules-of-thumb that we use so much, and more importantly, where the rules-of-thumb stop being applicable. Also, having used software developed and used by statisticians rather than by physicists, and joined their Internet mailing list community, I have had glimpses of another fascinating world.

Now that the thesis is finally finished, I hope that I have provided a new insight that will be an interesting and enjoyable addition to research literature.

Acknowledgments

Many thanks to my advisors Rolf Stabell and Sjur Refsdal.

A very special thanks goes to my good friend Miguel Carrión Álvarez, who has been of great help in this study. I have had many interesting discussions in physics and statistics with many of my friends and fellow students of these subjects, Stein Vidar Hagfors Haugan, Øyvind Saanum, Harald Fekjær, Jan-Erik Ovaldsen, Kathinka Dalland Evans, Inger Hansen, Øystein Olsen, Torstein Olsmo Sæbø, Erik Stensrud Marstein, Håkon Enger, Benedicte Selmer, Klaus Ade Johnstad and Henrik E. Nilsen.

I am happy for all the support I have had from my friends: I thank my girlfriend Monica Vibece Myrdal for her support and patience. I also thank Dagrun Vikhamar, Gro Fjerdumsmoen and Andreea Font for their support and advises.

Many thanks also to Helen Margaret Murray for the final proof-reading of the manuscript.

I would like to thank the authors of `R` and the `R` core team for their great skill and dedication in developing this software, and to all the members of the `R`-help mailing list, especially Brian D. Ripley, for their help.

Unrelated to the present thesis, I have also had the pleasure to work with Rudolph E. Schild on his QuOC-Around-the-Clock collaboration. I am grateful for this chance to do observational astronomy, in addition to this theoretical work.

I also acknowledge extensive use of NASA's Astrophysics Data System Abstract Service. I believe the ADS will make an important impact on the future of scientific publishing.

The last few chapters of this thesis was written at Bjørnholt located in the forest north of Oslo, where I rented a small hut to find peace and quiet to finish the work. I wish to thank the hosts for their hospitality.

Last, but not least, I my greatest thanks to my parents for getting me here, inciting an early interest in science, and supporting me.

Contents

List of Figures

List of Tables

1. Introduction

The present thesis is concerned with the Broad Emission Line Regions of Quasars. It is the intention to employ extra-galactic gravitational microlensing in the study.

Quasars are very remote and luminous objects in the Universe. They are thought to be the central regions of special galaxies. Special in the sense that they are particularly active, and active in the sense that they emit a lot of radiation. Often, the intensity of this radiation is variable in time.

A massive body deflects light rays. Under some conditions, a massive body may deflect light rays from a remote source just like a lens deflects light rays that pass through it, and therefore, the concept of "gravitational lensing" arises.

Gravitational lensing of quasars may, for example, occur when a massive galaxy between us and a quasar bends the light so that we observe multiple images of the same quasar. Since the galaxy contains stars or other compact objects, we may in addition have microlensing, which occurs when the stars or compact objects focuses the light on us, or spreads it away from us.

The Broad Emission Line Region of Quasars is thought to consist of discrete clouds surrounding a massive and highly luminous central engine. The number of clouds and the kinematics of this region are at present unknown.

So, why is this interesting? First, the study of the Broad Emission Line Region is currently a very hot topic, there is a great deal of controversy as to the structure of this region, and I hope that this new angle of view may contribute in that debate. Secondly, understanding the Broad Emission Line Region is important in order to understand the quasar as a whole, and microlensing represents a new way of determining quasar structure.

Finally, microlensing is a unique tool in exploring visible and dark matter in the intervening galaxy. To be able to use this tool with greater precision and more effectively, it is important to understand the emitting source.

1.1. Relation to previous studies

The first theoretical study of the Broad Emission Line Region using microlensing was by Nemiroff (1988). Later, Schneider & Wambsganss (1990) made an extensive investigation. These studies focused on a few physically motivated models, but the approach

prohibited very extensive statistical analysis. It was also concluded that certain models was not likely to produce detectable results. By making simplifying assumptions, I will study many of the models that previous studies found to be difficult, and thus the "worst-case scenario" part of the title. This investigation is not primarily concerned with making physically realistic assumptions, but rather making extensive statistical analysis feasible. As we shall see, a successful approach to the most difficult cases of previous studies can be made.

It should also be noted that empirical investigations, using the methods described in the previously mentioned papers, have been conducted, see e.g. Lewis et al. (1998). This paper concludes that microlensing has probably been observed.

A more detailed discussion of the relation between the present thesis and previous studies will be given in Section 4.7.

1.2. Structure of this thesis

Chapter 2, Theory opens with discussing in some detail the foundations of the gravitational lens theory, before I outline some important aspects of current views on quasars. Finally, I shall give a description of the statistical methodology that is used in this thesis.

Chapter 3, Distribution of Amplification deals with the question of how the probability of amplification of clouds are distributed. Both numerical and analytical distributions are discussed.

Chapter 4, Basic Spectral Line Profiles details the physical assumptions that are made, and how the spectral line profiles are made on the basis of these assumptions.

Chapter 5, Description of the test. The statistical test that is at the core of the analysis is detailed. Its capabilities in identifying amplified clouds are examined and its appropriateness for the problem at hand is assessed.

Chapter 6, Results gives a comprehensive overview of the results of the core analysis of this thesis.

Chapter 7, Variations addresses some variations of the core analysis.

Chapter 8, Discussion and Conclusions are given as the results are reviewed, before this text is ended with a summary of the important conclusions.

The Appendices contain the computer source code needed to reproduce the results of this study as well as tables of results that may be useful as reference.

Bibliography and an Index can be found at the end of the printed copy.

2. Theory

In this chapter, I shall outline the theory needed to understand the the issues addressed in this thesis. First, we shall study the basic physics of gravitational deflection of light and gravitational lensing. Then, we shall make an overview of quasar fundamentals before the chapter is concluded with a discussion of the statistical methodology that will be employed.

2.1. Gravitational deflection of light

The fundamental mechanism behind gravitational lensing is gravitational bending of light rays. The possibility that a massive object could bend light rays was discussed by Newton as early as 1704, and the Munich astronomer Soldner predicted in 1804 that a light ray passing near the limb of the sun would be deflected by 0.875 seconds of arc. It was never attempted to observe the predicted deflection.

Using his new theory of general relativity, Einstein predicted in 1915 that a ray of light would be bent when close to the limb of the sun by an angle

$$\hat{\alpha} = \frac{4GM_{\odot}}{c^2 R_{\odot}} = 1.75''. \tag{2.1}$$

which is twice that predicted by Soldner. Here G is the gravitational constant, c the speed of light and $M_{\odot}$ and $R_{\odot}$ the mass and radius of the sun, respectively. This result was confirmed by Eddington and his collaborators in 1919 during a solar eclipse. Presently, the result is confirmed to a precision better than 0.1% by VLBI measurements (see Lebach et al., 1995).

In general, a light ray passing a massive object with mass M by a minimum distance ξ is bent by the Einstein angle given by

$$\hat{\alpha} = \frac{4GM}{c^2 \xi} = \frac{2r_S}{\xi} \tag{2.2}$$

where we have inserted the Schwarzschild radius for the mass of the deflecting object:

$$r_S = \frac{2GM}{c^2}. \tag{2.3}$$

A more detailed treatment of this subject can be found in Refsdal & Surdej (1994).

As I shall discuss lensing at cosmological distances, I will now give a detailed description of the geometry of deflection of a light ray in a situation relevant to lensing. In Fig. 2.1, we see the geometry for one ray of light as it is deflected by a massive object (called a "lens"). It is assumed in the situations of interest to us that the extent of the lens along the line of sight is negligible compared to the distance traveled by the ray. It is also assumed that the mass distribution can be approximated by a point mass.

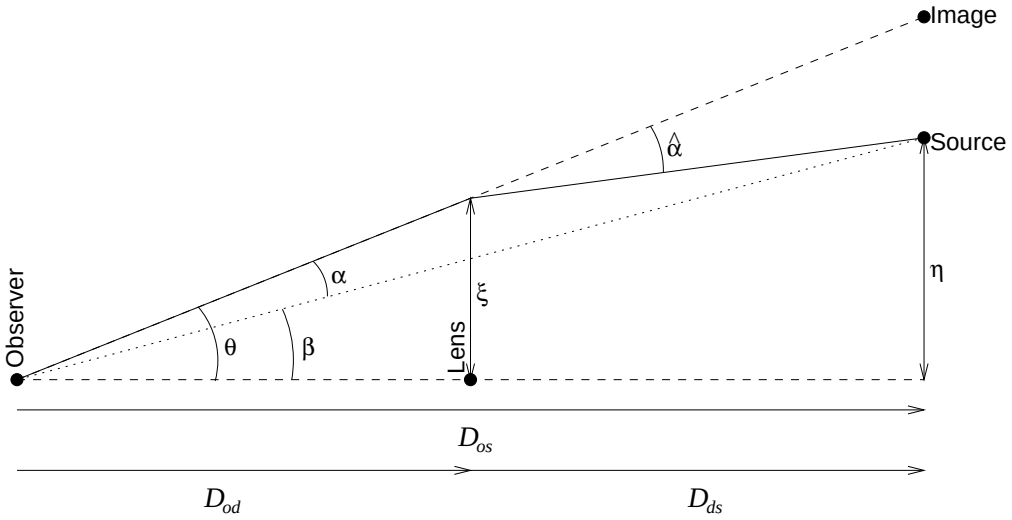

Figure 2.1.: *Geometric arrangement in which a ray is traced from a source at the right, passing a lensing body, to an observer on the left.*

It is conventional to use three planes, each perpendicular to the figure plane: The observer plane, the deflector plane and the source plane.

In this figure D_{os}, D_{od} and D_{ds} is used to denote the angular diameter distance (for definition, see e.g. Narlikar, 1993, Section 3.8) *from* the observer *to* the source, from the observer to the deflector and from the deflector to the source, respectively. It is important to emphasize that e.g. $D_{od} \neq D_{do}$ and $D_{os} \neq D_{od} + D_{ds}$, in fact $D_{od} = D_{do}/(1+z_d)$, where z_d is the cosmological redshift of the lens. Furthermore, θ is the observed angular separation between the lensing object and an image, β the separation between the lensing object and the source, $\hat{\alpha}$ the deflection angle induced by the lens, and α is the so-called reduced deflection angle, and is given as $\alpha = \frac{D_{ds}}{D_{os}}\hat{\alpha}$. ξ is used to denote the impact parameter, the minimum distance from the light ray to the lens, and η the separation between the true source position and the lens if projected into the source plane.

Assuming that angles are small, it is quite straightforward to show from simple geometry that

$$\eta = \xi \frac{D_{os}}{D_{od}} - \hat{\alpha} D_{ds}. \tag{2.4}$$

2.2. Gravitational lensing

Light rays from a distant source will follow different paths through spacetime, and be subject to gravitational fields of different shapes and strengths. In some fortunate cases, objects will be aligned in such a way that several of the paths ends up on earth, in which case multiple images may be observed.

While Einstein discussed the possibility of gravitational lenses, he eventually concluded that the phenomenon would probably remain unobserved. This conclusion was probably due to the fact that he only considered deflection by stellar masses. In 1937, Zwicky discussed many aspects of gravitational lensing, including lensing by galaxies, but noted to his surprise 20 years later that the phenomenon had not been observed. In the mid 1960s, Sjur Refsdal made many important contributions in showing how gravitational lensing can be used as an astrophysical tool. Again, see Refsdal & Surdej (1994) for more details. It was not until 1979 that a gravitationally lensed object was discovered, when the Twin Quasar was serendipitously discovered by Walsh et al. (1979).

We have up to now considered deflection of a single ray. We shall now consider the case of a thin lens deflecting multiple rays that pass the lens at different points in the lens plane, which is now said to be spanned by a vector $\vec{\xi}$. By "thin lens", we mean that all the mass is in the lens plane, which is not unreasonable since the distance from the lensing galaxy to the source and to the observer is vast compared to the relatively small extent of the galaxy itself. A superposition of deflection by point masses in the lens will make up the total deflection. Analogous to Eq. (2.2) one may write

$$\vec{\hat{\alpha}}(\vec{\xi}) = \frac{4G}{c^2}\iint \frac{\Sigma(\vec{\xi}')(\vec{\xi}-\vec{\xi}')}{|\vec{\xi}-\vec{\xi}'|^2}\mathrm{d}\vec{\xi}', \tag{2.5}$$

where the primed $\vec{\xi}'$ is a dummy variable that is used to denote the coordinates of the mass elements we integrate over, and we have introduced the surface mass density $\Sigma(\vec{\xi})$.

At this point, we may also introduce a dimensionless quantity, the *normalized* surface mass density

$$\kappa(\vec{\theta}) = \frac{\Sigma(D_{od}\vec{\theta})}{\Sigma_{cr}}, \tag{2.6}$$

where Σ_{cr} is the critical surface mass density for lensing, i.e. the surface mass density that will focus all the rays on the observer given as

$$\Sigma_{cr} = \frac{c^2}{4\pi G}\frac{D_{os}}{D_{ds}D_{od}}. \tag{2.7}$$

The most interesting special case to this study, is the case when $\kappa \sim 1$, which is the case when a lens may produce multiple images of the source, a situation discussed in the following section.

Because of magnification or rather (the observed property) amplification, gravitational lenses may also act as "natural magnifying glasses". Different parts of a ray bundle will experience gravitational forces of different magnitudes as it pass a lens, causing distortions in the original image.

Since no absorption or emission takes place when light is gravitationally deflected, the surface brightness is preserved. Additionally, the radiation is not frequency shifted beyond the cosmological redshift.

This means that the intensity I of an image is identical to that of the unlensed source (this holds also when using the specific intensity). The flux density F of an image is related to the intensity by

$$F = I\Delta\Omega, \tag{2.8}$$

where $\Delta\Omega$ is the solid angle subtended by the image in question.

It is then reasonable to define the amplification in terms of the ratio between the flux densities for a lensed image and its unlensed counterpart. Taking F as the flux density from a lensed image, I as the intensity of the same, and $\Delta\Omega$ as the solid angle subtended by the lensed image, and $\mathcal{F}$, i and $\Delta\omega$ as their respective unlensed counterparts, we may write the magnification factor μ as

$$\mu = \frac{F}{\mathcal{F}} = \frac{I\Delta\Omega}{i\Delta\omega}, \tag{2.9}$$

but since we've argued that $I = i$, this becomes simply:

$$\mu = \frac{\Delta\Omega}{\Delta\omega}. \tag{2.10}$$

We see that the magnification factor, in the following referred to mainly as the *amplification factor*, only depends on the ratio between the solid angles subtended by the lensed and unlensed images.

It can also be shown that the magnification can be obtained by finding the Jacobian determinant, see Section 5.2 of Schneider et al. (1992). With this approach, it is also possible to find the magnification of infinitesimal sources.

In conclusion, if the distortion makes the image bigger, it will also be brighter, as a whole.

2.2.1. Macrolensing

Illustrated in Fig. 2.2 is a typical case of gravitational macrolensing, a galaxy is situated between us and a remote quasar, and we observe several images. In fact, it can be proved (see e.g. Blandford & Narayan, 1986) that there will be an *odd* number of images, but also that we will almost always see just an *even* number of images because one image will be demagnified.

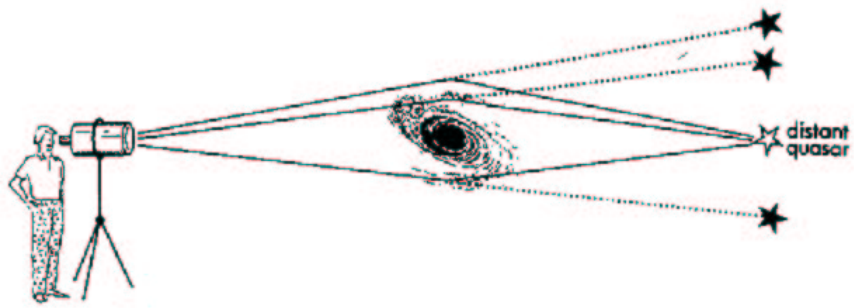

Figure 2.2.: *Observing through a telescope, one sees lensed images of a distant quasar at the positions where the extension of the path would leave them, as the light rays is deflected by the intervening massive galaxy. From Refsdal & Surdej (1994).*

In Fig. 2.3 I have included an image of a macrolensed object. In this unique object, known as Q2237+0305 or the "Einstein Cross", we see four images of a quasar at a redshift of 1.69, while the lensing galaxy is exceptionally close, at a redshift of 0.039 (Huchra et al., 1985). Note that in all but one presently known case, the source is at a redshift greater than 1, and consequently all quasars studied in the field of macrolensing belong to a class of very remote quasars. In the following I will use the term *multiply imaged quasar* to denote such a system.

Another interesting special case is when β in Fig. 2.1 is zero. It is then easy to see from the geometry that

$$\theta_0 = \hat{\alpha}\frac{D_{ds}}{D_{os}}, \tag{2.11}$$

where θ_0 is the angle that defines the *Einstein ring*. In the lens plane, the radius of this ring is called the *Einstein radius*, and by using Eq. (2.2) we find that it is given by

$$\xi_E = \theta_0 D_{od} = \sqrt{\frac{4GM}{c^2}\frac{D_{ds}D_{od}}{D_{os}}}. \tag{2.12}$$

If we introduce

$$D = \frac{D_{ds}D_{od}}{D_{os}} \tag{2.13}$$

and make use of the Schwarzschild radius Eq. (2.3), we may write this radius as

$$\xi_E = \sqrt{2r_S D}. \tag{2.14}$$

We shall also make extensive use of the Einstein radius as a natural length scale when projected into the source plane. It is given as

$$\eta_E = \xi_E\frac{D_{os}}{D_{od}} = \sqrt{2r_S D_{ds}}. \tag{2.15}$$

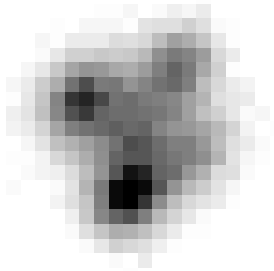

Figure 2.3.: *The so-called "Einstein Cross", a nearby galaxy lenses a remote quasar. We see four quasar images, three are easily seen, and the center of the galaxy. Image taken by the author using the Nordic Optical Telescope during A.O. Jaunsen's observing run on 1998-05-24.*

The Refsdal method for determining H_0

Quasars often show some variability in their luminosity. This property is useful for several applications, most notably for determining the Hubble parameter H_0. It was shown by Refsdal (1964) that if you have two images A and B, and θ_A, θ_B and θ_{AB} are used to denote the angular separation between the lens and the images and the two images respectively; and z_s and z_d the redshift of the source and lens respectively. Then

$$H_0 = \frac{z_d z_s \theta_{AB}(\theta_A - \theta_B)}{(z_s - z_d)\Delta t}, \tag{2.16}$$

where Δt is used to denote the time-delay between the two images. This equation holds only for small redshifts (that is, when Hubble's law holds), and for a spherical deflecting galaxy. The method was later generalized so as to hold for arbitrary deflectors in inhomogeneous universes by Kayser & Refsdal (1983). This method for determining H_0 has become known as "the Refsdal method".

In addition to intrinsic variation in the images, we might also see variations caused by a phenomenon known as *microlensing*, which is the subject for the next section.

2.2.2. Microlensing

Microlensing was first discussed by Chang & Refsdal (1979). They referred to the subject as "star disturbances". Later, Paczynski (1986) coined the term "microlensing", and gave a description of the phenomenon:

A real galaxy is made of stars and some continuously distributed mass: interstellar matter and possibly some exotic particles (dark matter). A detailed structure of a macroimage may depend on the masses and surface density of stars which split the macroimage into a large number of microimages, separated by some micro-arcsec. Even if we cannot resolve this structure, it may affect brightness of the macroimage and contribute to its variations in time.

Later that year, Kayser et al. (1986) extended this work in several directions, and thereby gave the study of microlensing a firm foundation. I shall discuss one of the most important equations, the normalized lens equation for microlensing shortly.

The first detection of a microlensing event was reported by Irwin et al. (1989).

In Fig. 2.4, a schematic presentation of microlensing and the technique of raytracing (which will be discussed shortly) is included. This figure is meant to be an aid to intuition in the following discussion.

The Normalized Lens Equation

In this section, I shall show how the normalized lens equation found in Kayser et al. (1986) can be derived. This equation is of great importance in the study of microlensing but it will not be used directly in this thesis. Note that I am not using the same notation as Kayser et al. (1986), as conventions have changed since this paper was written.

The total deflection angle of both macrolensing and microlensing was given by Kayser et al. (1986) as:

$$\vec{\hat{\alpha}} = \vec{\hat{\alpha}}_0 + \frac{4G}{c^2} M \vec{S}(\vec{\xi}) - \frac{\kappa_c}{D}\vec{\xi} + \frac{1}{D}\begin{pmatrix} \gamma' & 0 \\ 0 & -\gamma' \end{pmatrix}\vec{\xi}, \tag{2.17}$$

with

$$\vec{S}(\vec{\xi}) = \sum_i \frac{\vec{\xi}_i - \vec{\xi}}{|\vec{\xi}_i - \vec{\xi}|^2} \tag{2.18}$$

and D defined in Eq. (2.13) and κ_c (the dimensionless surface density of continuously distributed matter) is in units of Σ_{cr} as it was defined in Eq. (2.7).

It is important to emphasize that this equation gives the *total* deflection a light ray experiences, both the deflection caused by the deflecting galaxy as a whole, and microlenses. In our final result, however, we are interested in tracing the rays from one particular image. This may be confusing, as the geometry of Fig. 2.1 is used at two different levels, to describe both the macrodeflection and the microdeflection, and so is most of our notation. In Fig. 2.4, the top panel shows the deflection caused by the macrolens, and the additional deflection caused by microlenses are shown in the middle panel. Also, the origin of $\vec{\xi}$ is marked, making its meaning clearer: In this context it is the impact parameter relative to a microlens, not the center of gravity of the macrodeflector (galaxy). Likewise, $\vec{\eta}$ is the separation between the true source position and the

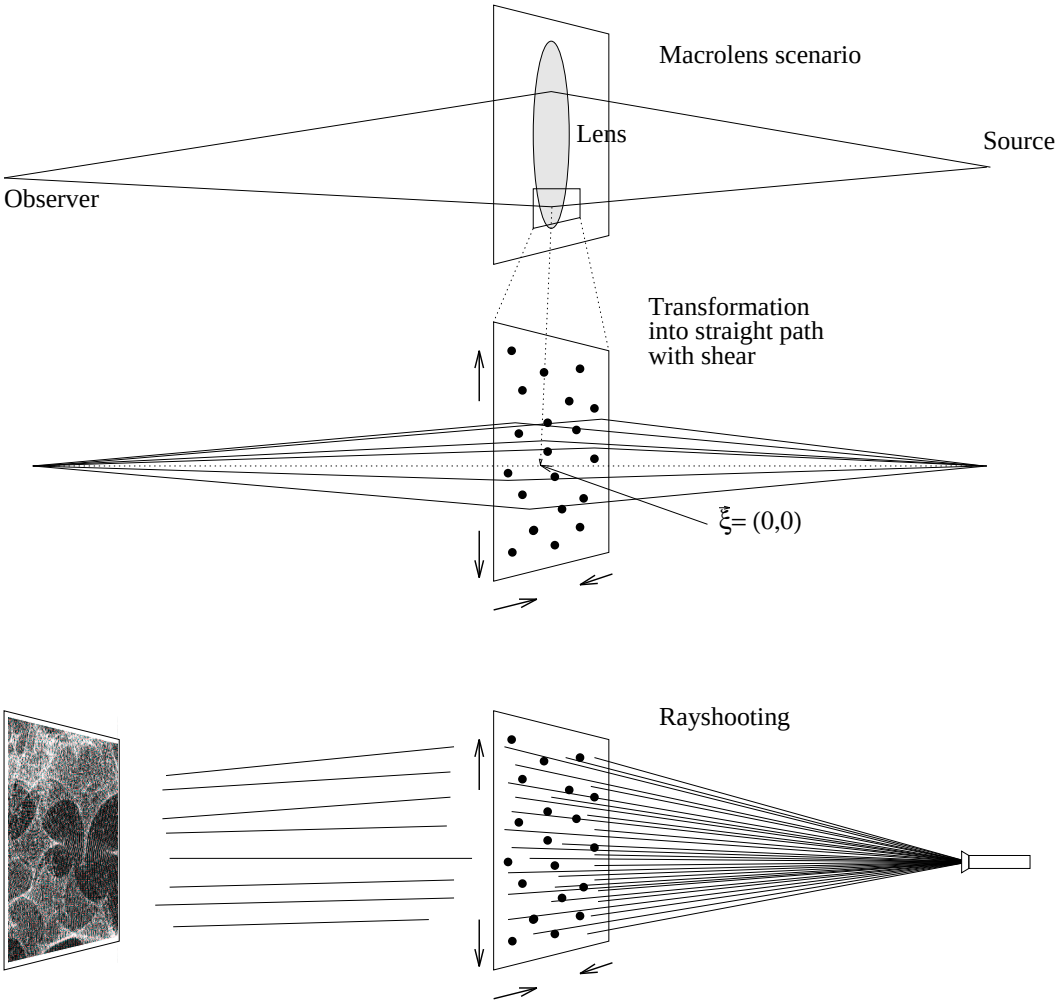

Figure 2.4.: *Illustrating the principle of microlensing and raytracing. In the upper figure, the macrodeflection is shown. The figure in the middle shows a starfield from the deflecting galaxy, and the lower figure illustrates raytracing with a torch. In practice, the rays will not be shot radially from a source, but will be shot in parallel into the starfield. From Haugan (1994).*

origin of $\vec{\xi}$ if projected into the source plane. The macrodeflection is $\vec{\hat{\alpha}}_0$, and once that is realized, the rest of the derivation is concerned with microlensing only.

The second term of Eq. (2.17) is very much analogous to Eq. (2.5), but instead of integrating over an infinite number of small mass elements, we sum over a finite number of stars. Note here that the deflection induced by stars is proportional to $\vec{\xi}^{-1}$.

The third term considers deflection from continuously distributed matter with a constant surface mass density. We note that the deflection angle from such a deflector is

$$\vec{\hat{\alpha}} = \frac{4\pi G\Sigma}{c^2}\vec{\xi}. \tag{2.19}$$

If the units of κ_c and the simple relation $\xi = \theta D_{od}$ (which is easily seen from the geometry in Fig. 2.1) are used, it is easy to see that we get the third term of Eq. (2.17). The negative sign makes a ray bundle contract.

The last term of Eq. (2.17) is known as the "shear term". Shear is induced in the lens by the macrodeflector, and influences the patterns in the raymap by braking the symmetry. Imagine a circular bundle of rays being deflected by e.g. a galaxy nucleus as depicted in Fig. 2.5. All rays will be deflected by an amount proportional to the inverse of the distance, and we see from the matrix of Eq. (2.17) that the bundle will be compressed as much vertically as it is stretched horizontally. In Fig. 2.5, the vertical component of the forces is depicted along with the forces themselves, and this is done to illustrate the same point: The difference between the length of the horizontal arrows is the same as the sum of the length of the vertical arrows.

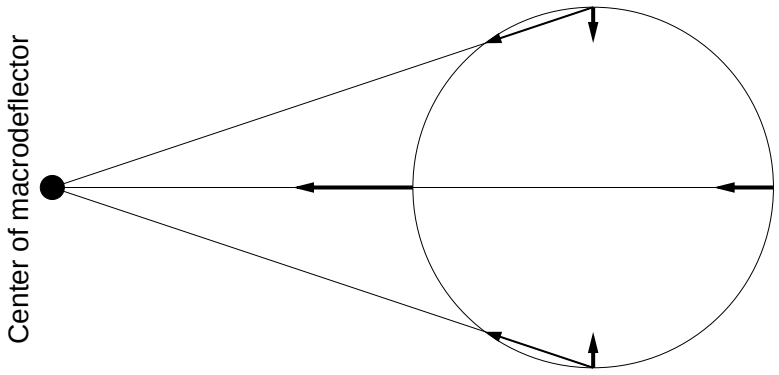

Figure 2.5.: *Shear. A circular bundle of rays pass by a macrodeflector; the ray bundle is deformed by differential gravitational forces.*

Kayser et al. (1986) introduce the ray tracing equation on the form

$$\vec{\eta} = (\vec{\hat{\alpha}} - \vec{\hat{\alpha}}_0)D_{ds} + \vec{\xi}\frac{D_{os}}{D_{od}}. \tag{2.20}$$

This equation relates the true source position relative to the projected origin of the microlens $\vec{\eta}$ with the deflection angles $\vec{\hat{\alpha}}$ and $\vec{\hat{\alpha}}_0$ from the microlens and the macrolens

respectively and the impact parameter of a ray traced past the microlens, as well as the angular size distances between the three involved planes. Note that this equation reduces to Eq. (2.4) in the case where there is only macrodeflection.

If Eq. (2.17) is inserted into the above equation, we obtain

$$\vec{\eta} = \frac{4G}{c^2} M\vec{S}(\vec{\xi}) D_{ds} - \frac{\kappa_c}{D} D_{ds}\vec{\xi} + \frac{1}{D} \begin{pmatrix} \gamma' & 0 \\ 0 & -\gamma' \end{pmatrix} D_{ds}\vec{\xi} + \vec{\xi}\frac{D_{os}}{D_{od}}. \tag{2.21}$$

To arrive at Eq. (4) of Kayser et al. (1986), we need the normalized length unit in the deflector plane introduced by Paczynski (1986)

$$\xi_0 = \left(\frac{4GDM}{c^2|1-\kappa_c|} \right)^{1/2} = \frac{\xi_E}{\sqrt{|1-\kappa_c|}}, \tag{2.22}$$

with ξ_E from Eq. (2.12), and then, the corresponding normalized length unit in the source plane:

$$R_E = \xi_0(1-\kappa_c)\frac{D_{os}}{D_{od}}. \tag{2.23}$$

This is the Einstein radius for microlensing in the source plane and will be used in so many contexts, I have found it convenient to denote it by R_E. It is also often referred to as η_0.

The $\vec{\xi}$'s and the $\vec{\eta}$'s in Eq. (2.21) all have their physical dimensions, and so do the length units in Eq. (2.22) and Eq. (2.23). To get a dimensionless quantity, define:

$$\vec{\xi}' = \frac{\vec{\xi}}{\xi_0} \quad \text{and} \quad \vec{\eta}' = \frac{\vec{\eta}}{R_E} = \frac{\vec{\eta}}{\xi_0(1-\kappa_c)}\frac{D_{od}}{D_{os}}.$$

Then, substitute these for the $\vec{\xi}$'s and the $\vec{\eta}$ in Eq. (2.21):

$$\begin{aligned} \vec{\eta}'\xi_0(1-\kappa_c)\frac{D_{os}}{D_{od}} &= \frac{4G}{c^2\xi_0} M\vec{S}(\vec{\xi}') D_{ds} - \frac{\kappa_c}{D} D_{ds}\vec{\xi}'\xi_0 \\ &+ \frac{1}{D} \begin{pmatrix} \gamma' & 0 \\ 0 & -\gamma' \end{pmatrix} D_{ds}\vec{\xi}'\xi_0 + \vec{\xi}'\xi_0\frac{D_{os}}{D_{od}}. \end{aligned} \tag{2.24}$$

We rearrange to find that

$$\begin{aligned} \vec{\eta}' &= \frac{4GM}{c^2\xi_0^2(1-\kappa_c)}\frac{D_{ds}D_{od}}{D_{os}}\vec{S}(\vec{\xi}') - \frac{\kappa_c}{1-\kappa_c}\frac{1}{D}\frac{D_{ds}D_{od}}{D_{os}}\vec{\xi}' \\ &+ \frac{1}{(1-\kappa_c)D}\frac{D_{ds}D_{od}}{D_{os}} \begin{pmatrix} \gamma' & 0 \\ 0 & -\gamma' \end{pmatrix} \vec{\xi}' + \frac{\vec{\xi}'}{1-\kappa_c}. \end{aligned} \tag{2.25}$$

If we insert ξ_0 given in Eq. (2.22) in the first term, and recall D from Eq. (2.13), then

$$\vec{\eta}' = \frac{|1-\kappa_c|}{1-\kappa_c}\vec{S}(\vec{\xi}') - \frac{\kappa_c}{1-\kappa_c}\vec{\xi}' + \frac{1}{1-\kappa_c} \begin{pmatrix} \gamma' & 0 \\ 0 & -\gamma' \end{pmatrix} \vec{\xi}' + \frac{\vec{\xi}'}{1-\kappa_c}. \tag{2.26}$$

To obtain Eq. (4) of Kayser et al. (1986) we must observe that the second and the last term of Eq. (2.26) becomes

$$-\frac{\kappa_c}{1-\kappa_c}\vec{\xi}' + \frac{\vec{\xi}'}{1-\kappa_c} = \frac{1}{1-\kappa_c}\vec{\xi}'(-\kappa_c+1) = \vec{\xi}' \tag{2.27}$$

and also

$$\text{sign}(1-\kappa_c) = \frac{|1-\kappa_c|}{1-\kappa_c}\left(= \frac{1-\kappa_c}{|1-\kappa_c|}\right). \tag{2.28}$$

Rename $\vec{\xi}'$ and $\vec{\eta}'$ to $\vec{\xi}$ and $\vec{\eta}$. We have arrived at Eq. (4) of Kayser et al. (1986):

$$\vec{\eta} = \text{sign}(1-\kappa_c)\vec{S}(\vec{\xi}) + \vec{\xi} + \frac{1}{1-\kappa_c}\begin{pmatrix} \gamma' & 0 \\ 0 & -\gamma' \end{pmatrix}\vec{\xi}. \tag{2.29}$$

To go from here to Eq. (6) of Kayser et al. (1986) is pretty straight-forward, we just need two convenient quantities, first the normalized surface mass density *in stars*

$$\kappa = \frac{\kappa_s}{1-\kappa_c}, \tag{2.30}$$

and the normalized shear parameter

$$\gamma = \frac{\gamma'}{1-\kappa_c}. \tag{2.31}$$

Using an identity matrix when factorizing $\vec{\xi}$, we get

$$\begin{aligned} \vec{\eta} &= \text{sign}(\frac{\kappa_s}{\kappa})\vec{S}(\vec{\xi}) + \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \gamma & 0 \\ 0 & -\gamma \end{pmatrix}\right)\vec{\xi} & (2.32)\\ &= \text{sign}(\frac{\kappa_s}{\kappa})\vec{S}(\vec{\xi}) + \begin{pmatrix} 1+\gamma & 0 \\ 0 & 1-\gamma \end{pmatrix}\vec{\xi}. & (2.33) \end{aligned}$$

κ_s and κ_c will always be positive, and therefore $\text{sign}(\frac{\kappa_s}{\kappa}) = \text{sign}(\kappa)$ must be valid. Thus, the normalized lens equation is

$$\vec{\eta} = \text{sign}(\kappa)\vec{S}(\vec{\xi}) + \begin{pmatrix} 1+\gamma & 0 \\ 0 & 1-\gamma \end{pmatrix}\vec{\xi}. \tag{2.34}$$

Model classification

In Fig. 2.6, the shape of caustics are indicated for different microlens models, in the case where the microlens is a single star, in addition to some continuously distributed matter and the gravitational field from the galaxy, the so-called “Chang-Refsdal lens”. The

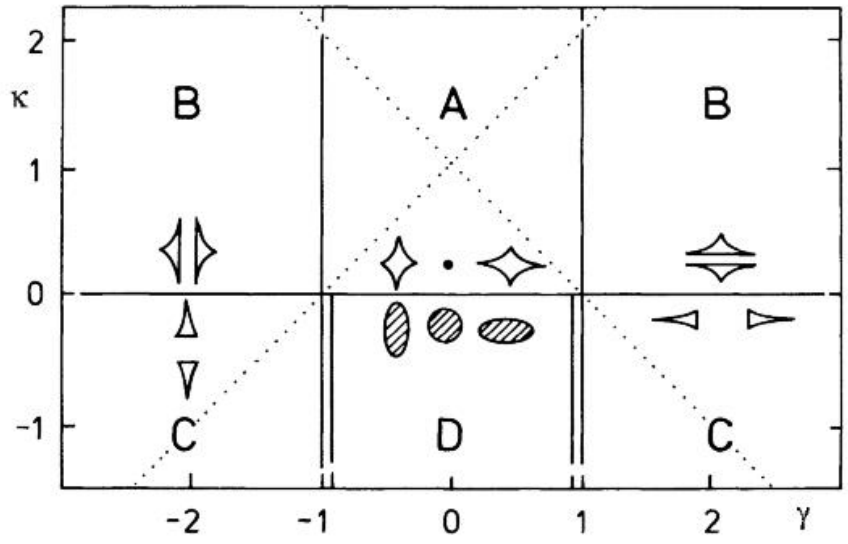

Figure 2.6.: *Classification of microlens models. The critical curves mapped to the source plane (caustics) for different values of γ and κ, in the case of a Chang-Refsdal lens. From Kayser et al. (1986).*

sign of γ is seen to dictate the orientation of the caustic model curve, and its magnitude decides how much its form is stretched. It also decides whether or not the caustic is stretched so much that it splits up in two closed model curves.

The sign of κ determines whether the rays are under-focused or over-focused. It is easy to get confused here, because in principle, a gravitational lens is infinite in extent (because a gravitational field is infinite), so one might think that some rays will never be over-focused. Recall, however, Eq. (2.30), which states that also the continuously distributed matter contributes to the focusing of the gravitational lens. Consequently, there will be over-focusing when the contribution from continuously distributed matter is large enough to make κ negative. Because only rays passing through the continuously distributed matter will be focused, it is implicitly assumed here that the source is physically smaller than the macrodeflector, as projected into the source plane. It is obvious that this assumption holds, however, as the typical macrodeflector is orders of magnitude larger than any source, since the largest source in question is an active nucleus of a galaxy, while the macrodeflector is a whole galaxy.

For zero shear, the shape of the caustic is always circular or a point. However, in the case of over-focusing, that is $\kappa < 0$, the model curve is a dark circular area with very few photons. With some shear we get a dark ellipse or model D in Fig. 2.6. With $\kappa > 0$ and some shear, we get the frequently seen "diamond shaped caustic", we have A-type models. For large shear terms, it is seen that the caustic model curve splits up in two.

Caustics are interesting because high amplification may occur when a source is on,

or close to, a caustic. This will be discussed later.

Lightcurve production by making raymaps

Even more interesting than the caustic model curves of Fig. 2.6 are so-called *caustic networks*. These appear when rays are shot through a field with a large number of scattered stars (raytracing). The patterns created will resemble those above, but there will be several, they will be twisted due to the gravity of the surrounding starfield, complex patterns may appear where stars are close to each other and so on.

Normally, *backward* raytracing is practiced, that is, rays are traced from the observer, through the lens and onto the source plane.

The purpose of a raymap is to show how many rays will hit a particular area. We may normalize the raymap, so that the value of each pixel represents the amplification μ of a source with the size of a pixel. Since a value of 1 means an unamplified source, this is done by dividing the value of each pixel by the number of rays expected from an unlensed source (or one may use the average of a *very* large raymap).

For a circular source larger than the size of a pixel, we find the number of rays within the source radius and divide it by the number of rays that would have been found within the same area in the unlensed case. The resulting number is the amplification for that source. We may also use the normalized raymap, in which case the amplification is the average of the values in every pixel that falls within the source. It is also possible to use sources with internal structure. We may for example use a source that is fainter at the edges than in the center, and use, for example, a normalized Gaussian to reflect this. I will, however, use mainly pixel sized sources.

We are interested in the amplification at *different* positions in a raymap. Due to the fact that the stars, the galaxy, the source, and the observer are in motion relative to each other, a source will move along a "track" relative to the pattern in the raymap with a velocity corresponding to the transverse component of the resulting velocity of all the relative motion. Recording the amplification of a source with its center at different positions along the track will give a *lightcurve*.

The normalized source radius is

$$R = \frac{R_{phys}}{R_E}, \tag{2.35}$$

with R_{phys} as the physical (proper) radius of the source and R_E is the Einstein radius of the microlens as defined in Eq. (2.23).

Producing a lightcurve is usually done by taking a source with a specified radius and move it across a raymap. The amplification of the source is computed for as many positions as desired, or as allowed by the pixel size. One may compute the amplification for at most every pixel along the track. Plotting for every position then yields the lightcurve.

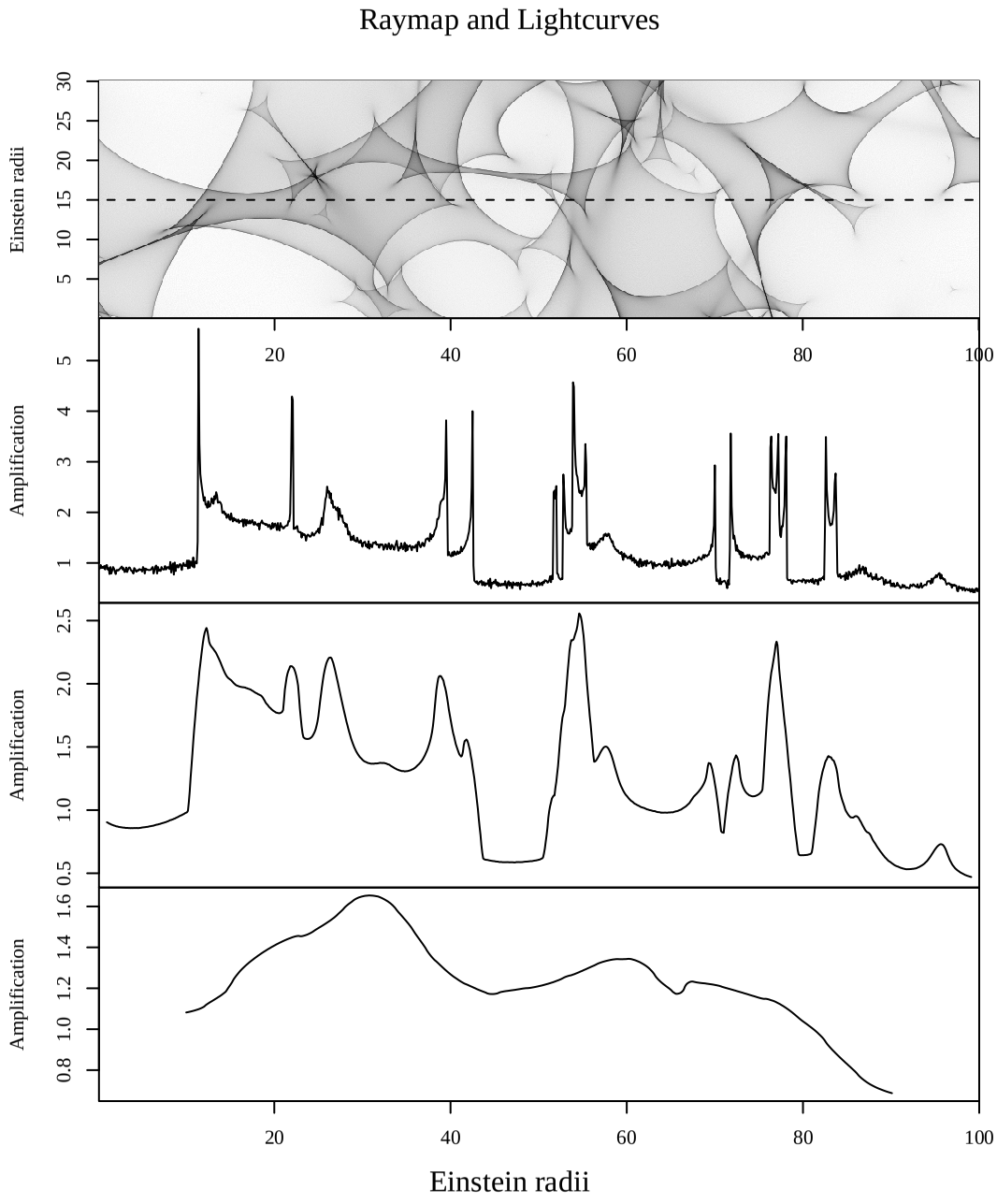

Figure 2.7.: *Raymap (top panel) and lightcurves for a small ($R = 0.01R_E$), medium ($R = 0.1R_E$) and large source ($R = 1R_E$) as they appear if a source follows the track shown by the dashed line. $\kappa = 0.5$, $\gamma = 0$. A circular source with uniform brightness was used.*

In Fig. 2.7 we see that with small sources, there may be very strong effects. This is due to the fact that within the source radius, the number of rays counted when on a caustic will far exceed that of the unlensed case. For large sources on the other hand, points far from the caustic will keep the average amplification down, and thus smooth the curve. This process can be thought of as a running average filter.

As we shall see, the possible strong effects for small sources is a property that is essential in this study: It is why we may have hopes of observing the effects of individual clouds, due to their suspected small size (of order 1 AU).

Lightcurve production by contouring

The contouring method was described by Lewis et al. (1993) and Witt (1993).

It involves drawing an infinitely long straight line across the source plane. The line is then mapped to the lens plane as viewed by the observer. The resulting curves are called "image curves". The problem of constructing a lightcurve is then a matter of searching for the image curves, and then follow them.

Mathematically, let $f(\vec{\xi})$ be the function that maps a point in the source plane to a point in the lens plane. The image of the source line (corresponding to $f(\vec{\xi}) = \eta_y =$ const) at a given separation between the planes is then a contour of this function (thus the name "contouring method").

It is shown in Lewis et al. (1993) that the image curves consists of one infinite line across the lens plane and a number of closed curves, each crossing at least one star.

The search for the infinite line can be done by starting on the outskirts of the starfield, where the image line is close to the position it would have been in if no stars were present, and we can therefore expect to find this line quite fast. The closed curves are found equally fast by starting at the stars in the field.

With all the images of the source line known, the image curves are followed, and the magnification is extracted. Combining the magnification for all the images for each corresponding point in the source plane yields the lightcurve.

However, as shown by Haugan (1999), this method is only exact for idealized point sources. For extended sources of *any* size, there are systematic deviations.

There is another point that will be of greater importance to us than direct lightcurve production: Since the line in the source plane is infinite it divides the source plane into two distinct regions. Therefore, the line's mapping to the lens plane must necessarily divide the lens plane into distinct regions as well, projected on either side of the image curves. This will prove useful.

Haugan's `cmap`

Stein Vidar Hagfors Haugan has developed a `c`-program that implements the *hybrid* method for raymap and lightcurve production. This method uses both the contouring method and the raytracing method.

Haugan's program, outlined in Haugan (1996), is very flexible, with a number of command-line options. Among them are the normalized surface mass density in stars κ, the shear γ, the direction of the shear. One may also use a rectangular star field, the code contains proper corrections for such use. I will be using only circular starfields, and will not be discussing issues that are related to non-circular starfields.

One may also specify the size of the starfield, or let the program calculate it using the formula found by Schneider & Weiss (1987) to ensure that all points in the target area receives 99 % of the diffuse flux.

The number of stars is calculated based on the surface mass density, and the stars are distributed randomly over the field.

Since the mapping of one line in the source plane produces image curves that divide the lens plane into regions, the mapping of *two* such lines will delimit regions that are between the two corresponding image curves in the lens plane (in addition to regions above and below). Note that these regions consist of the region enclosed between the infinite lines that pass across the plane, as well as the regions associated with the closed curves found by stars. Adding the image curves corresponding to another two infinite lines, orthogonal to the first two, the regions in the lens plane corresponding to a rectangular source can be found as the intersection of the regions between the two pairs of image curves.

Instead of finding two full sets of curves corresponding to two sets of orthogonal *infinite* lines, only two full image curves (corresponding to two parallel lines) are found. Of the orthogonal lines, only those segments connecting the two first lines are calculated (the starting points of all such segments can be found on the first two lines).

This produces closed loops corresponding to images of a finite, rectangular source plane region, enclosed by the two sets of (imaginary) orthogonal lines in the source plane. With this, all the images of a finite, rectangular source plane region have been found, and conventional backward raytracing is then performed by tracing rays through the closed loops back to the source plane. Thus, only rays falling inside the "source rectangle" are computed in the raytracing part of the operation.

Time scales and model degeneracy

It is natural to define a time scale based on the length scale in the source plane, that is, the Einstein radius, and the transverse velocity of the source relative to the critical curves, with time measured in the observer's time frame. This velocity was given by Kayser et al. (1986) as

$$\vec{v}_T = \frac{1}{1+z_s}\vec{v}_s + \frac{1}{1+z_d}\frac{D_{os}}{D_{od}}\vec{v}_d + \frac{1}{1+z_d}\frac{D_{ds}}{D_{od}}\vec{v}_o, \tag{2.36}$$

where $\vec{v}_s$, $\vec{v}_d$ and $\vec{v}_o$ are the transverse peculiar velocities of the source, the star field and the observer respectively, in their own plane. The natural time scale becomes

$$\Delta t = \frac{\Delta\xi}{\vec{v}_T}. \tag{2.37}$$

It can be shown that the same lightcurves may be produced if we scale the mass of the compact objects by a factor of f^2 and adjust the physical positions, the source size and the transverse velocities by a factor of f. It is currently not possible to measure all these quantities at the same time, only relative quantities can be inferred from the observations of lightcurves. This unfortunate fact is the degeneracy problem of microlensing, and great caution is required when trying to constrain parameters from microlensing.

To break this degeneracy, one may use three telescopes far apart in space. This was first shown by Kayser et al. (1986) and Grieger et al. (1986). Haugan (2002) studies the feasibility of such a project.

High Amplification Events

There seems to be no consensus as to an accurate definition of a high amplification event (hereafter HAE) in the literature. What is meant by HAE is however that an image is seen to vary rapidly in brightness, as much as one magnitude is not uncommon. In Fig. 2.7, we see several HAEs for small sources. What causes an HAE is not unambiguous, however, as it is only connected to the amplification recorded in each case. In most cases in Fig. 2.7, an HAE is connected with a caustic crossing. In general this may or may not be the case. If a small source crosses just outside the end of a caustic (called a “cusp”), we may also see very strong amplification. By comparing the lightcurve for the medium sized source with the raymap, one can see a good example of this. It is seen that the second peak stems from a caustic crossing, whereas the third peak (which is in this case stronger than the second peak) stems from a cusp. It would be very difficult to distinguish an event stemming from a cusp from a single caustic crossing. On the other hand, if there is a dens caustic network (which may occur for $\kappa \approx 1$), the amplification may always be high, and caustic crossings will not produce a significant relative amplification.

The envisioned clouds around the quasar core will be discussed extensively in this thesis. These clouds are very small and if a cloud is highly amplified by being on or near a caustic, a single cloud may contribute significantly to the total luminosity on one specific frequency of a quasar spectrum. This is the core subject of this thesis.

2.3. Quasars

2.3.1. Properties and history

The first quasars were discovered in the early 1960s. They were first given the name "Quasi-stellar radio source" (QSRS), because they were discovered at radio wavelengths, and looked like stellar objects on optical photographs. Later, similar objects with little radio emission were observed mainly in the visual wave band, and so the abbreviation QSO for "quasi-stellar object" eventually appeared. The two terms have commonly been used to denote objects that send out relatively intensive radiation in radio (i.e. are "radio-loud") and those that do not (i.e. are "radio-quiet"), respectively.

It is now clear that these objects are nuclei of extremely active galaxies and it is therefore suggested that the terms QSRS and QSO are abandoned, and that "quasar" is considered a new word. The terms Radio-loud and Radio-quiet should be used to distinguish between the two classes if desired.

Quasars are a part of the broader class of galaxies with Active Galactic Nuclei, commonly abbreviated "AGN". The other classes are Seyfert Galaxies, Blazars, Radio Galaxies and other related objects, like LINERs. These classes can be subdivided into several narrower subclasses.

When AGNs are classified, the characteristics one looks for are:

- if they are point sources,
- if they have emission at the same order of magnitude in visual, infrared and X-ray light,
- if they have broad lines in their spectra,
- if they have narrow lines in their spectra,
- if there is considerable radio emission from the object (about 10^{-3} of bolometric luminosity),
- if the object is variable, and
- its polarization characteristics.

The properties that characterize quasars are firstly that they are point sources, and secondly, they do have broad and narrow lines. There are some differences between radio-quiet and radio-loud quasars: The former show weak variability, while in the latter class, some show strong variability, and the same is the case for polarization, where the criterion is that it should be at least a few percent linear polarized light coming from the source. More details about the classification of other AGNs can be found in the excellent table on pg. 20 of Krolik (1999). It should be noted that Seyfert 1 galaxies differ only from radio-quiet quasars in that in the former class, some objects show

strong variability. It is thus rather difficult to make a clear criterion for what is a quasar and what is a Seyfert, as noted by Robson (1996) on pg. 85.

2.3.2. Structure, energy generation and transport

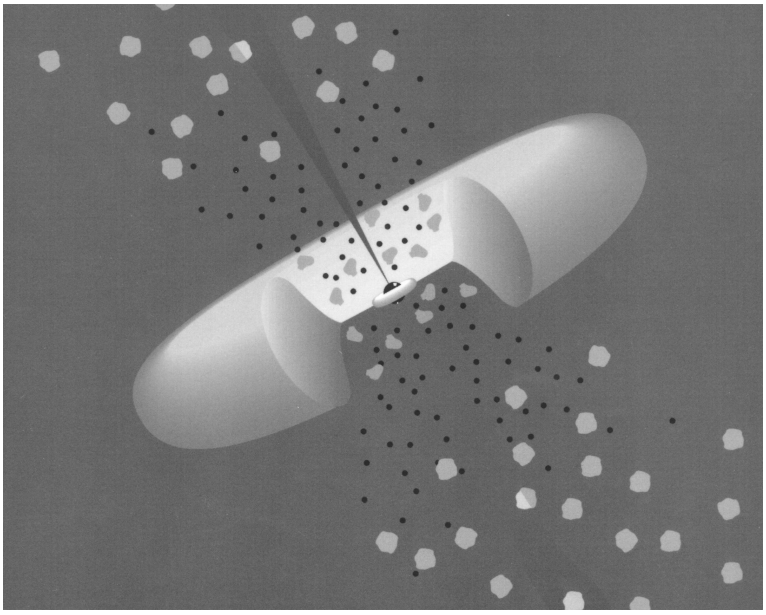

Figure 2.8.: *Artist's conception of a Quasar. Part of the molecular torus has been removed so that the central black hole is visible. From the black hole, a jet, and around it the accretion disk. We see the broad-line clouds (close to the center), and the clouds of the narrow line region further out. The black dots depict a region with hot electrons. Thanks to Meg Urry for an explanation of the figure and for permission to use it.*

There is still some controversy as to what is the fundamental engine of quasars, but most researchers have settled for a central black hole model. The most discussed alternatives are starburst models. The essence of these latter models is that a large number of supernovae are thought to produce the observed emission.

As depicted in Fig. 2.8, the central engine consists of a supermassive (that is mass $> 10^6 M_\odot$) black hole in the very center. Around the black hole, a disk of matter has gathered, and friction caused by the viscous forces in the gas causes it to fall towards the black hole, and also to heat up. This disk is therefore called an accretion disk. The accretion disk is believed to be of size of the order of 10^{15} meters and radiate dominantly thermal radiation. At about 10^{11} meters from the center, we are in a region of intense EUV, X-ray and γ radiation, that extends down towards the black hole.

Around the black hole and the accretion disk, there is a large molecular torus, on a 100 pc-scale (10^{18} m). This is important to us because in unification models (models that aim at unifying all AGNs, mainly by saying it is essentially the same object viewed from different angles), this torus may hide parts or all of the Broad Emission Line Region, which is of primary interest in this study. This torus is heated on the inside by photo-ionization. This is very nicely illustrated on pp. 334 – 335 of Robson (1996).

There are emitting regions of little importance to our study on rather large scales, such as the Narrow Line Region and the Extended Narrow Line Region. These are also depicted in Fig. 2.8, but I shall not discuss these regions, but end this section by discussing the most important region to us, the Broad Emission Line Region.

The Broad Emission Line Region

This region, hereafter referred to as the BLR, is generally believed to consist of discrete clouds, and is heated by photo-ionization. The photons responsible for the heating are coming from the accretion disk.

There are many good arguments for the BLR being heated by photo-ionization. I shall briefly list some of the arguments that can be found in Krolik (1999) starting on page 319:

1. Fluctuations in strengths of some lines have been seen to follow fluctuations in the continuum flux with a time-delay, and that delay is roughly consistent with what has been predicted by models.

2. Photo-ionization is preferred to collisional ionization because the inferred temperature of the clouds is not high enough to overcome the ionization potentials of some of the species seen. On the other hand, the continuum provides a reasonable amount of sufficiently energetic photons to ionize atoms.

3. Photo-ionization models that has been constructed are able to reproduce the relative line strengths very well.

4. The Equivalent Width of lines are relatively constant over a wide range of luminosities. If the total flux from the line increases, we will see an increase in both the central part and the wings of the lines, so that the equivalent width will not change significantly. This suggests that the energy release from the quasar central engine is responsible for the energy release in lines.

Each cloud contributes to the total line profile with a narrow line, Doppler-shifted from the cloud rest-frame frequency because of the velocity of the cloud along our line of sight.

There is at this time no consensus in the literature as to how the velocity field of clouds are. Done & Krolik (1996) evaluates a number of models on the basis of observations of NGC 5548. They prefer models that have substantial random motion and

net infall (more matter falling towards the center than matter flowing out). They also reject a few models, among them, all outflow models. This is disputed by Bottorff et al. (1997), who are working on hydromagnetic wind models (see Emmering et al., 1992). In these models, clouds are accelerated away from the nucleus radiatively and centrifugally in a hydromagnetic wind. In a recent review, Sulentic et al. (2000) argue, however, that there is evidence of outflow at the "boundaries" of BLR.

It is nevertheless clear, that no simple model based on a single dominating mechanism is consistent with observations. Presumably, a realistic model must take into account gravitation, drag forces, winds and so on.

It is beyond the scope of this thesis to distinguish between these models. I shall therefore make some very rough approximations when it comes to cloud kinematics, and the cloud velocity model with be discussed in Section 4.1.3.

Finally, one should be aware that there is considerable controversy as to the very existence of discrete clouds. It has been argued by Arav et al. (1998) based on high-resolution Keck data, that the broad emission lines in NGC 4151 could not have been produced by discrete emitters.

2.3.3. Spectra

It is expected that the best spectroscopic instruments available must be used for observations if one has to have any hope of detecting microlensing of individual clouds.

A detecting device will have discrete *bins*, so that all photons having frequency in an interval $f + \Delta f$ will hit a specific bin, where Δf is the width of the interval.

These bins will be of great importance in this study. A broad emission line will hopefully cover a number of bins, on the order of hundred is expected. Each bin will contain a count of how many photons have hit this bin.

Information about instruments are available from the World Wide Web: There are information about Keck's High Resolution Echelle Spectrometer (HIRES) available at
<URL:http://www2.keck.hawaii.edu:3636/realpublic/inst/hires/hires.html>
and VLT's UV-Visual Echelle Spectrograph (UVES) at
<URL:http://www.eso.org/instruments/uves/>.

2.4. Statistics

2.4.1. Distributions

The basic properties of a Distribution Function $F_X(x)$ (using X as a generic random variable and x as its realization) are:

1. The Distribution Function is nondecreasing.
2. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.

3. Since the Distribution Function is monotonic it can only have "jump" discontinuities. "Jump" discontinuities means that finite probability is found at a point.

(see e.g. Dudewicz & Mishra, 1988, pp. 70).

The Distribution Function is thus often called the *cumulative* distribution function, describing the probability that a random variable X takes on a value smaller or equal to x. The second property above is the normalization condition.

The discontinuities mentioned in the third property will only arise with discrete distributions. A continuous random variable will have an absolutely continuous Distribution Function. For an example of a Distribution Function with jumps, consider the Poisson distribution, which will be much used in this thesis. An example of the Distribution Function of the Poisson distribution can be seen in Fig. 2.9.

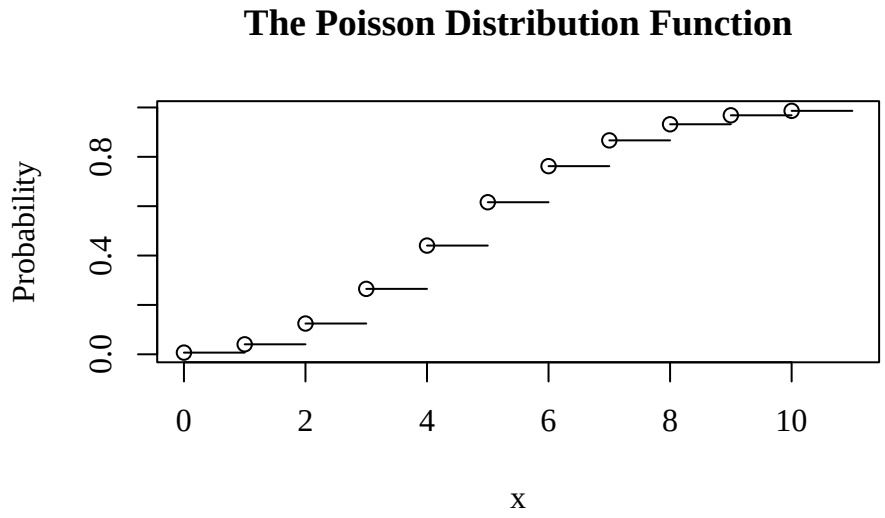

Figure 2.9.: *The Poisson Distribution Function with a parameter of 5. The graph shows the probability of drawing a number smaller than x.*

We shall be working mostly with continuous random variables, however, and they must obey the following condition as well:

$$F_X\left(x \to x_0^-\right) = F_X\left(x \to x_0^+\right), \tag{2.38}$$

for all x_0.

For continuous variables, another important concept is the *Probability Density Function*. The Probability Density Function of a distribution can be found by simply differentiating the Distribution Function. The Probability Density Function is useful in that

one may obtain the probability that a random variable X takes on a value x in an interval around x by integrating the function over that interval. In the discrete case, the corresponding function is called simply the "Probability Function".

For a plot of a Probability Density Function, refer to Fig. 2.10, which depicts a χ^2-distribution – another distribution that is important in this thesis.

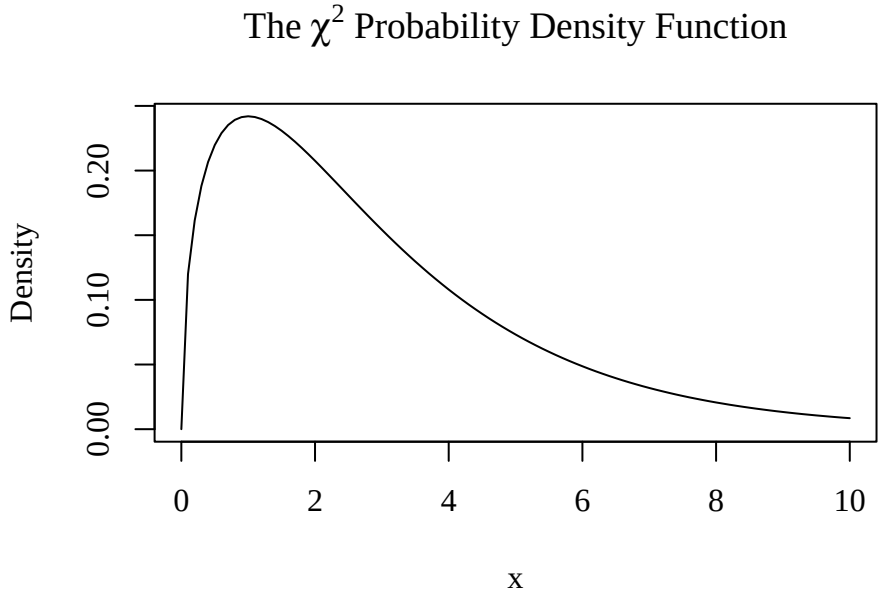

Figure 2.10.: *The χ^2 Probability Density Function with 3 degrees of freedom.*

Finally, an important function is the inverse of the Distribution Function, namely the Quantile Function. In this thesis, its utility is mainly connected to drawing numbers from distributions. When the Quantile Function is found, a standard way to draw numbers from a distribution, is to draw numbers in the interval between 0 and 1 (something that most statistical software has support for). Then, the Quantile Function is evaluated for the number drawn (see, e.g. Dudewicz & Mishra, 1988, pp. 180 for details). I use this method in Section 3.2, with some modifications in some cases.

There exists a number of distributions that are used in many applications, that are named and that are well described in the literature. They are often described by a small number of parameters. For instance, the "Normal distribution" (also called the "Gauss distribution"), is characterized by its mean (often denoted μ) and its standard deviation (often denoted σ). I shall give more details about two other important distributions below:

The Poisson Distribution

We have already seen a Distribution Function of the Poisson distribution in Fig. 2.9. This distribution is important in many astronomical applications, especially in so called "counting processes", for reasons that will soon become apparent.

First, let us define the Poisson distribution by stating its Probability Function:

$$\mathrm{P}(X = x) = \frac{e^{-\vartheta}\vartheta^x}{x!} \tag{2.39}$$

where x is a non-negative integer and ϑ is the *Poisson parameter*.

One very useful and important property of the Poisson distribution is that both its average and its variance is equal to its parameter, ϑ.

Consider the following example: During some time interval, an event (such as detection of photons) may or may not occur. Under three important assumptions, one may use a Poisson distribution in this context:

No simultaneous events: Two events may not occur at the same time.

Independence: The number of times the event occurs in any interval is independent of the number of times it occurs in any other disjoint time interval.

Time homogeneity: The average number of times the event happens in an interval does not change with time, and is equal to ϑ. It is, however, possible to modify the Poisson parameter in such a way that a variable can be said to be Poisson distributed also in a time-dependent case.

Under these assumptions, the Poisson distributed random variable may be thought of as a count of events, and therefore the term "counting process" is used.

When the last assumption is known to hold, it is very useful. For one thing, it means that if an event is known to occur on average ϑ times in a time interval, then if the length of the interval is modified by a factor of k, then, we will have a new Poisson distributed random variable with a parameter $k\vartheta$.

Events occurring in space may be thought of in analogous terms: If we find on average ϑ items in an area, then, modifying the area with a factor k will also yield a new Poisson distributed random variable with a parameter $k\vartheta$.

The χ^2-distribution

We have seen a Probability Density Function of a χ^2-distribution plotted in Fig. 2.10. The Probability Density Function of the ${\chi^2}_n$-distribution with n degrees of freedom can be written as

$$f_X(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & \text{for } 0 \leq x < \infty \\ 0, & \text{otherwise,} \end{cases} \tag{2.40}$$

where Γ is Euler's Gamma function.

In Section 4.5.1, Babu & Feigelson (1996) describe an important application of the χ^2-distribution in astronomy: A goodness-of-fit test for binned data. Binned data is very common in astronomy: A spectrograph is the most obvious example of binning, but one can also think of a CCD as using 2-dimensional binning, where each pixel is a bin. Finally, classifications, such as the classification of AGNs into radio-loud/radio-quiet quasars, type I and II Seyfert galaxies, etc., can be thought of as binning.

Rephrasing the discussion from Babu & Feigelson (1996), let us consider an experiment consisting of N identical and independent trials, where the outcome of each trial falls into one of k bins. We have derived a model prior to the experiment that states that the probability of the single trial having an outcome that falls in bin i is p_i. We wish to see if the observed data $n_1, \ldots, n_k$ supports our model. n_i is the count of outcomes in each bin, and $n_1 + \ldots + n_k = N$. Also, $p_1 + \ldots + p_k = 1$.

Then for large N,

$$\sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i} \tag{2.41}$$

is approximately χ^2-distributed with $k-1$ degrees of freedom, if $p_1, \ldots, p_k$ are the true values.

There are many other examples of use of the χ^2-distribution, of which I shall not go into detail. For example, it is often used in regression or to make inferences about the variance of a variable with normal errors.

This will be discussed further in Section 2.4.3.

2.4.2. Hypothesis testing

Hypothesis testing is a way of answering questions such as "is it justified from available data to say that the mean of X is x_0?" and "are we justified from the available data to say that lensing is detected?"

In essence, hypothesis testing is about quantifying two mutually exclusive conjectures and then use probability theory to decide between them.

These two conjectures are normally referred to as the *null hypothesis* H_0 and the *alternative hypothesis* H_1. It is conventional in statistical methodology to use H_0 to denote the established conjecture and H_1 to denote the new competing hypothesis. Then, we may reject H_0 if the evidence mounted against it by H_1 is sufficient, but one should not refer to H_1 as "accepted" if that happens. (see e.g. Bhattacharyya & Johnson, 1977, Chapter 6 for discussions).

There are two types of errors that can be made, either failure to reject H_0 when H_1 is true (e.g. we say there is no lensing, when there is) or rejecting H_0 when H_0 was true (e.g. we say there is lensing when there isn't). Conventionally, the latter error (often labeled "Type I Error") has been considered more serious than the former (often labeled "Type II Error"), and therefore, one usually names a level of acceptable probability

of committing this error, typically 1, 5 or 10 percent, known as the *significance level* (usually just the "level") of the test. The level is often denoted α. Given a test with a fixed level, the probability of making a Type II Error is called the "power" of the test, hereafter denoted ϕ. When I use the term "detection" about the result of a hypothesis test, it only means that H_0 is rejected on a α level.

Generically, we formulate the test in terms of some characteristic for either hypothesis, for example: "are the means different", "are the variances different", "are the numbers in one case consistently higher than the other"? There are two main classes of methods: parametric methods and non-parametric methods. In the parametric methods, one makes an assumption on the distribution of the variables in the H_0 and H_1 case, then goes on to estimate the parameters of these distributions.

In non-parametric tests, we make no assumption about the distribution of the variables involved. Since for many astronomical purposes, the distribution of the variables are unknown, non-parametric methods are of great importance. Non-parametric tests use many different concepts and approaches, and I shall not go into details with more than one such method, below.

Common for both types are the use of the *test statistic*. The test statistic is a condensation of all the information in all the gathered data. It is itself a random variable that is computed from the other random variables we observe. The test statistic is used to determine when H_0 is to be rejected. The set of values for which H_0 is rejected is called the rejection region.

Take, for example, one of the simplest parametric tests, the one-sample t-test: We have a data set consisting of a random sample $X_1, X_2, \ldots, X_n$ assumed to be normally distributed $N(\mu, \sigma)$ and wish to test whether its estimated mean $\hat{\mu}$ is equal to some figure μ_0. The standard deviation σ is not known. Then

$$\hat{\Theta} = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{n}} \tag{2.42}$$

is said to have Student's t-distribution with $n-1$ degrees of freedom, given that $\hat{\sigma}$ is the estimated sample standard deviation.

We must reject H_0 at an α level if

$$\hat{\Theta} \leq -t_{\alpha/2,n-1} \qquad \text{or} \qquad \hat{\Theta} \geq t_{\alpha/2,n-1}, \tag{2.43}$$

where $t_{\alpha/2,n-1}$ is computed by statistical software or looked up in statistical tables for the t-distribution.

In the above equations, $\hat{\Theta}$ is the test statistic, and, as we see, it is a function of $\hat{\mu}$ and $\hat{\sigma}$, both of which need to be estimated from data. In other situations, the expression for the test statistic may be more complex, but the principles are the same: The decision to reject H_0 is based on the value the test statistic takes.

In fact, the expression in (2.41) is also a test statistic, where the value computed, when compared with the relevant ${\chi^2}_{k-1}$-distribution, is used to decide whether or not to reject H_0. In that example, H_0 was that the modeled probabilities did fit the data.

Tests themselves should be evaluated before they are trusted. Some tests are text-book examples, there are many situations where it is quite obvious that a well described text-book example test will work perfectly well. There are however, many cases where that is not obvious, as such examples always involve assumptions that should be checked before the test is applied. One cannot assume *a priori* that any generic test will work correctly, and to study the performance of a test, one must calculate the distribution of the test statistics in both the H_0 and in the H_1 case. When it is impossible or impractical to do this analytically, we may attempt to simulate both parameters.

When analytical or simulated distributions of test statistics are available, we may not only see if the test works, we may also probe its performance and if desired compare it to other tests.

To make an assessment of the applicability of a test we proceed as follows: We first do a simulation of H_0 and then test to see if H_0 is rejected. This may sound self-contradictory, but testing to see if the test rejects H_0 when we input a simulation of H_0 is exactly the point. If it does, there is something wrong with the test. However, because of statistical fluctuations, we must accept that the test erroneously rejects H_0, it is merely an example of committing "Type I Error", as described above. The probability of committing this error, must however, be held at a reasonable level. To address this, we may do a large number of simulations, and for each simulation, we perform the test. If the mean of the test statistic converge to the value expected under the null hypothesis, then we can conclude that the test on average does not reject H_0 falsely, which is the most basic criterion for a hypothesis test. We should also seek to estimate the density of the test statistics (by, for example, plotting them in a histogram).

We continue to examine the alternative hypothesis. We should follow the same procedure as above; compute a large number of simulations of H_1 and for each simulation, perform the test. For each test, we record the test statistic. Again, the probability density of test statistics should be estimated. The mean of the test statistic should be substantially larger than for H_0. If the test statistics for H_1 are not larger than for H_0, there are two possibilities: Either, there are really no differences between H_1 and H_0 (thus, H_0 stands, H_1 fails), or there is a difference, but the test is unable to establish it (we have used a poor test). If we know that we have done simulations of two different cases, then we should of course conclude that the test is inadequate for this purpose.

It is pertinent to ask how great the difference between the recorded test statistics of H_0 and H_1 must be. We stated above that it is conventional to use a level of 1, 5, or 10%, and the level being the probability of making a "Type I Error". However, fixing the level in this way allows no way to control the power (the probability of "Type II Error"). Fortunately, since we are able to simulate both hypotheses, we need not set the level at such an arbitrary point: We may make sure we find the optimal level.

There are many ways of defining what is "optimal", however. Most strategies involve comparing the higher end of the density of test statistics for H_0 to the lower end of the density of the test statistics for H_1. There are several ways to compare these, to list a few:

1. Equal probability densities (histogram columns are of the same height),
2. equal slopes of the probability densities,
3. equal tails (the areas of the tails in the density are equal).

I shall use the last (partly because in some cases studied later, the histograms I use are non-overlapping).

It is possible to show (Álvarez, 2000, private communication) that if we have the same number of statistics for both H_0 and H_1, this amounts to merging the two data sets and take the median:

Given two Distribution Functions $F_X(x)$ and $G_X(x)$, the probability left of a point x is $F_X(x)$ and $G_X(x)$ and on the right of the same point is $1-F_X(x)$ and $1-G_X(x)$. The higher tail of the Distribution Function for H_0 can be written as $1-F_X(x)$ and the lower tail of the Distribution Function of H_1 can be written as $G_X(x)$. Setting these two be equal is

$$1-F_X(x) = G_X(x), \tag{2.44}$$

or

$$F_X(x) + G_X(x) = 1. \tag{2.45}$$

By definition, the median m is a number that satisfies

$$\mathrm{P}(X \leq m) = \frac{1}{2}, \tag{2.46}$$

$$\mathrm{P}(X > m) = \frac{1}{2}. \tag{2.47}$$

A Distribution Function can per definition be written as $F_X(x) = \mathrm{P}(X \leq x)$. We see that $x = m$ and therefore that the median of the two data sets can be used as an optimal level.

Thus, in the case where both the null hypothesis and the alternative hypothesis can be simulated, an optimal *rejection threshold* may be found by performing a large, but equal, number of simulations for each hypothesis. Perform the test for each of the simulations, and compute the median of all the test statistics.

Let us proceed to the issue of making an assessment of the quality of the test. This can be done qualitatively by examining the density estimates we produced when we recorded the test statistics. As long as the density estimates (e.g. the histograms) of H_0 and H_1 have a finite overlap, there is a finite chance of making either mistake described above. Simply put, there is no way we can tell from the value itself if the test statistic comes from H_0 or H_1. The integrated density below the rejection threshold is the probability of making a Type II Error, the integrated density above the rejection threshold is the probability of making a Type I Error.

Thus, the greater the separation between the two density estimates, the better the test is. Further, if it is found that the overlap is so large that the rejection threshold computed above corresponds to a level larger than the conventional 1, 5 or 10%, then

one should label the test "unreliable", and I shall refer to this level as the *"reliability level"* and denote it by α_r.

Different quantities are cited when one reports characteristics of the test that has been committed. Usually, the level has been chosen (to 1, 5 or 10%), so that is normally reported. The test statistic may also be reported, but one may see that more often, a "p-value" is cited instead of the test statistic or even the level. As the p-value is the upper tail of the distribution limited by the test statistic, the p-value is the smallest level (α) for which H_0 could still be rejected. Because it is a probability, it is easier to compare with other tests, and knowledge about the distribution from which it is found is not needed to make sense from it. However, since it is also monotonically dependent on the test statistic, there may be cases where doing the transformation from test statistic to p-value is unnecessary. Cases where very high test statistics are transformed to p-values are archetypical examples where a loss-of-precision problem occur in most software. I have verified that the software I use does not have this problem when used as documented. One may nevertheless want to report the test statistic rather than the p-value in such cases.

The p-value is only a function of the statistics for the lensed case. The level, on the other hand, is a function of the statistics of both the lensed and unlensed cases, and in addition has not had an arbitrary threshold imposed. Therefore, I will use the level as the main characteristic when discussing the tests.

The rejection threshold will be elaborated on in Section 5.4.

2.4.3. Contingency Tables

The typical use of contingency tables as taught in elementary textbooks in statistics (e.g. Bhattacharyya & Johnson, 1977, Section 13.4) is to study whether certain levels of one characteristic tend to be associated with some levels of a different characteristic.

In astronomy, one good application is for certain types of binned data. On page 27 we briefly discussed a goodness-of-fit test for binned data. The contingency table is an extension of this formalism. Instead of comparing a vector and a given distribution, we compare different vectors, each coming from different categories we wish to compare. The data is organized in a table, by counting the number of samples that falls in each bin or category.

In mathematical terms, we write the contribution to the test statistic Θ_{ij} from each row i and column j as

$$\Theta_{ij} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \tag{2.48}$$

where O_{ij} is the observed number of counts in a bin (category), E_{ij} is the expected counts for that bin under the null hypothesis. Normally, we do not know what to expect, so the E_{ij}'s would have to be estimated. Let us say we have k rows and l columns. A

conventional estimator is

$$\hat{E}_{ij} = \frac{\left(\sum_i^k O_{ij}\right) \cdot \left(\sum_j^l O_{ij}\right)}{\sum_{\mathrm{all}ij} O_{ij}}. \tag{2.49}$$

For each row and column, Θ_{ij} is computed, and a large value of Θ_{ij} indicates that the number of counts in that bin is somewhat exceptional.

Furthermore, we wish to obtain a test statistic for the hypothesis test, and this test statistic is simply the sum over all bins of Θ_{ij}, $\Theta = \sum_i^k \sum_j^l \Theta_{ij}$. For a high number of observations, Θ has a χ^2-distribution with $(k-1)(l-1)$ degrees of freedom (see e.g. Dudewicz & Mishra, 1988), so that we will reject the null hypothesis on an α level if

$$\Theta > \chi^2_{(k-1)(l-1),1-\alpha}. \tag{2.50}$$

2.4.4. Histograms

The histogram is a very useful way of illustrating how frequently certain values of a random variable occur.

Quite simply, one divides the range of observations into bins, counts the number of observations with a value in a bin, and plots the number of observations (the frequency) as boxes, as a function of values. As a rule of thumb, $1 + \log_2$(number of observations) equidistant bins are used.

The number and the position of the breaks may, however, be chosen anyway one desires (as long as one is aware that spurious patterns may arise), and I shall for some applications use a large number of bins. In these cases, however, there is a large number of observations.

One may also see in some histograms that the density is plotted, instead of the frequency. In such cases, one may obtain the probability of observing a value in that bin by multiplying the density by the bin width. If I use the density, I shall also quote the bin width in the figure text.

One may wish to see double-logarithmic histograms of some data sets in this thesis. However, double-logarithmic histograms are deprecated in statistical practice. One reason is that one must avoid having any bins with zero counts, and one would have to carefully choose the breaks between bins. This may again bring out spurious patterns in the histograms, that are further amplified by the log-log transformation.

2.4.5. Software

I have used almost exclusively the Free Software system named `R`. This software has been described by Ihaka & Gentleman (1996), and is primarily a language and environment for statistical computing and graphics. While some features are still lacking, for astronomers notably in digital image processing, it is in rapid development by users

from all over the world. The system is an implementation of the `S` system. Another implementation of the `S` system that may be more familiar to many users of statistical software is `S-plus`.

`R` contains many powerful features, including some object-orientation, making it very easy to write structured and mathematically intuitive code. The fact that it is free software implies that the source code is available, and thus it is possible to seek understanding of all the internal workings in the software. While I have not sought a deeper understanding of the software, I have examined parts of the code quite carefully and on occasions submitted patches, and have been acknowledged as a contributor. As it is free software it also means that the code is under constant scrutiny by its users, and so undergoes a persistent but informal peer-review. This comes in addition to the fact that much of the code has been through formal peer-review in statistical and computer scientific journals.

Sometimes, I use the term “variable” and “object” interchangeably in this thesis. Readers familiar with programming languages will appreciate the meaning of “variable”, but there really is no such thing in the `R` language, there are only “objects”. For example, an integer variable is an object with type integer. `R` purists will object to my use of the term “variables”, while readers unfamiliar with `R` or object oriented programming may find the use of “objects” confusing. I have tried to find a middle ground to make the text clear to all.

3. Distribution of Amplification

For many purposes in this thesis, I need a function which represents the probability of having a cloud amplified by a certain amount.

There are many ways of making such a function or distribution, I will discuss the most important ones.

3.1. Numerical amplification distributions

The raymap will serve as a basis for creating numerical amplification distributions. A raymap is produced by tracing rays from the source, through the lens to the observer plane, or vice versa (backward raytracing), thereby producing a raymap which indicates how many rays have hit a specific area. See Section 2.2.2 for further details. Then, the raymaps may be convolved with a source of some size and normalized so that every pixel in the raymap has the value of the amplification at that point. Then, a histogram for amplification with many bins can be used as a Probability Function.

I have used code developed by Stein Vidar Hagfors Haugan (see Section 2.2.2) to produce the raymaps. Care must be taken so that the final Probability Function for each parameter set accurately represents the distribution of amplification for that parameter set, and this turned out to be a non-trivial task. The problem is to ensure that we have a "typical" star field.

In calls to Haugan's `cmap`, one may, among other parameters, give the normalized surface mass density in stars, the size of the star field in Einstein radii, the size of the field in pixels, and a "quality factor" which is used to determine the number of rays shot in every pixel. Roughly, this number corresponds to the average number of hits in each pixel. Since no ray should be lost or gained by lensing, the average amplification is unity[1] and thus the amplification in any pixel is given by the value of the pixel divided by this quality factor.

Since the clouds studied are small, we may conveniently choose the pixel size in Einstein radii to be of the order of the expected size of a cloud. The pixel size in Einstein radii is simply the size of the field in Einstein radii (an input parameter to

[1] In practice, some rays will be lost due to the finite size of the shooting range, so the mean amplification is slightly smaller than one.

cmap), divided by the field size in pixels, (another input parameter to cmap). This saves us the trouble of convolving the raymap with a source.

When using a pixel size that is orders of magnitude smaller than an Einstein radius, as clouds are expected to be, the problem arises that the size of a "typical" star field is rather large. Wambsganss et al. (1990) found a lot of structure even at the $100R_E$-scale for the Einstein Cross. Because the Probability Function obtained will be used to draw an amplification at an arbitrary point, it is critical that the raymap used to generate the Probability Function represents all possible amplifications and their relative abundance.

With a pixel size of for example $0.01R_E$ and a needed $100R_E$ field, the raymap would be several hundred megabytes, rather cumbersome to read into a computer's memory. I have solved this problem by computing a large number of smaller raymaps with different seeds rather than a single large one. In addition, one may convince oneself that the total area is sufficiently large by studying how the mean amplification converges to the expected value, namely slightly below unity.

In the ensuing chapters of this thesis, I will refer to the amplification distributions made in this section in several contexts. Since they are characterized by the pixel size, I will make reference to these distributions by citing their pixel size denoted by the symbol P_d.

3.1.1. Pixel size of $P_d = 0.01R_E$

I have computed 249 raymaps with a pixel size of $P_d = 0.01R_E$ with normalized surface mass density in stars $\kappa = 0.5$ of size $2 \times 2R_E$ and thus 200×200 pixels, and a quality factor of 200. I load one by one raymap, divide each pixel value by the quality factor and find the mean amplification of the raymaps loaded. By this method I can see when the average amplification converges towards unity with reasonable precision. This is illustrated in Fig. 3.1. We see that it takes more than 200 raymaps of this size for the mean amplification to converge with three significant digits.

With the convergence of the mean of the 249 raymaps in Fig. 3.1, one can be quite confident that a sufficiently large starfield has been used. It should be noted that such plots are merely indicative of convergence, they may differ substantially in appearance, depending simply on in which order the raymaps were loaded. One might use some statistical test, I have not done that, but I have loaded the raymaps in arbitrary order. We may then study the histogram of amplifications. In Fig. 3.2, I have included a histogram of the amplification for the same raymaps as above.

We note that the highest amplification in this example is 26.6, the lowest at 0.205. We also note with interest that the profile does not descend monotonically from the peak to the end. I have not made any investigation as to the origin of the features seen in this histogram, but note that Lewis & Irwin (1995) reported similar features.

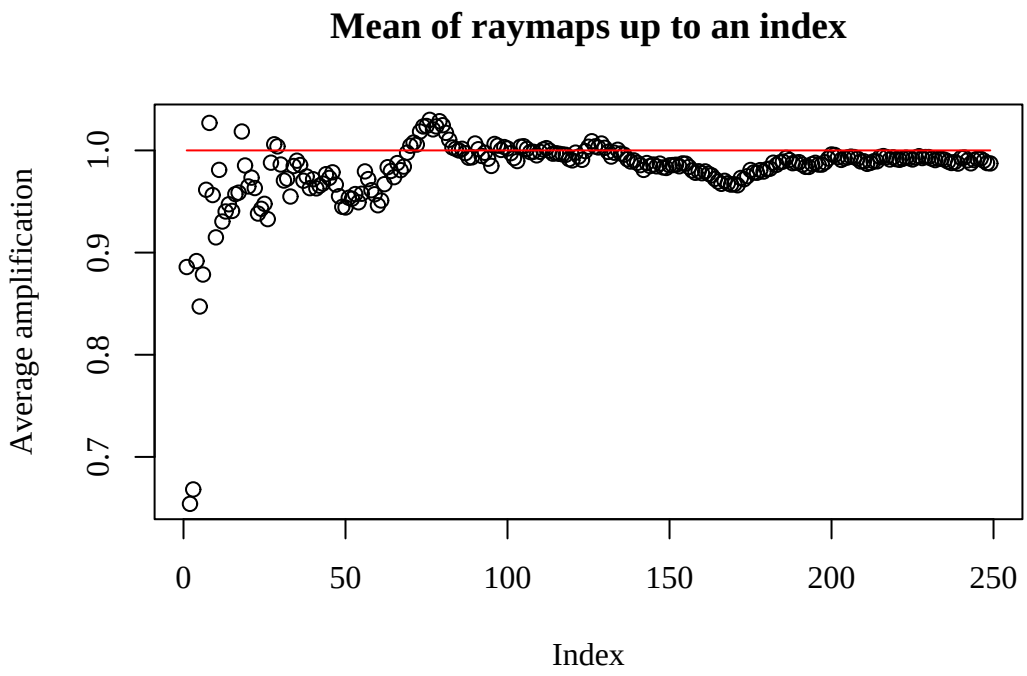

Figure 3.1.: *The average amplification of all raymaps up to an index. The average converges towards a value slightly lower than one (indicated by the red line). In this example, the pixel size is $P_d = 0.01R_E$ and the normalized surface mass density in stars is 0.5.*

3.1.2. Pixel size of $P_d = 0.0025R_E$

I have also computed 1001 raymaps with a pixel size of $P_d = 0.0025R_E$ with the same normalized surface mass density in stars, $\kappa = 0.5$, of size $0.5 \times 0.5R_E$, the same 200×200 pixels and quality factor of 200.

Following the same procedure as in the previous case, we first study how the average amplification evolves with an increasing number of raymaps. In Fig. 3.3 it is seen that the average amplification levels out rather well when approaching thousand, but the final value in this plot is 0.974. This means that a slightly higher number of raymaps might be diserable, but the only consequence is that the distribution is not known quite as precise as the one for $P_d = 0.01R_E$.

Now, to use the histograms in figures 3.2 and 3.4 as Probability Functions, they only need to be normalized, which is a trivial operation.

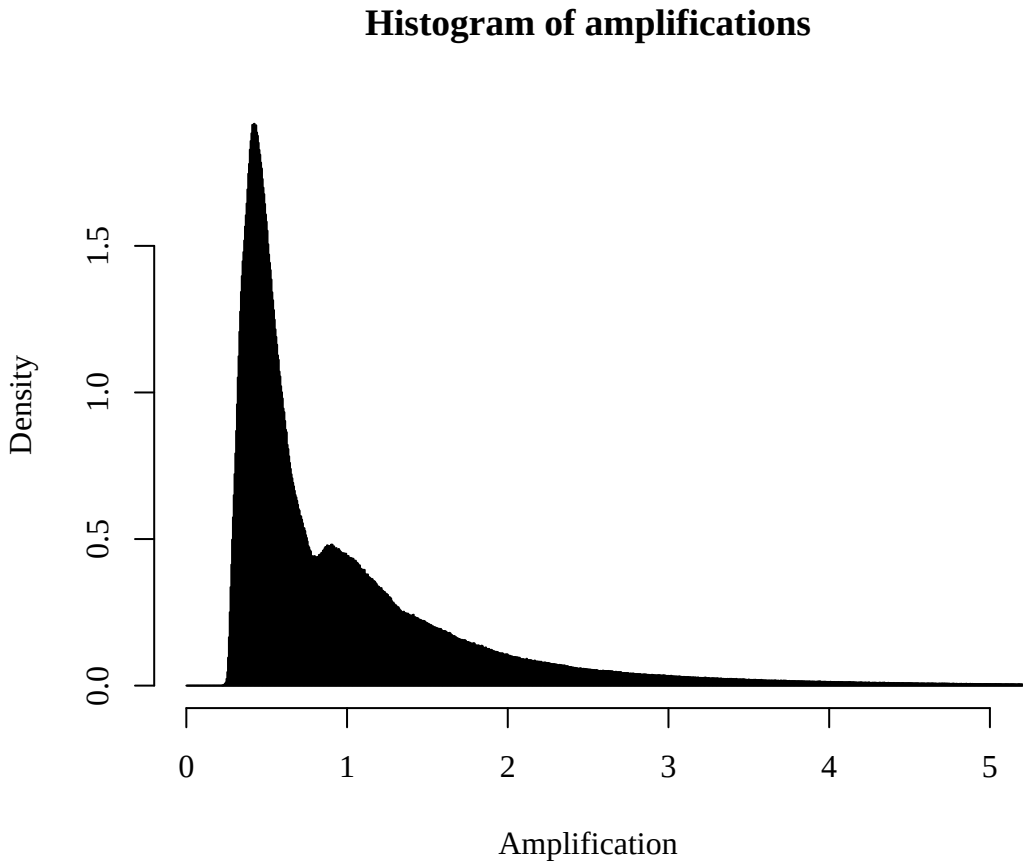

Figure 3.2.: *Histogram of amplifications. The pixel size is $P_d = 0.01 R_E$. The figure only shows amplifications from 0 to 5; the density above 5 is very small. There are a total of 9 960 000 data points in this case, and the amplification bin width is 0.005.*

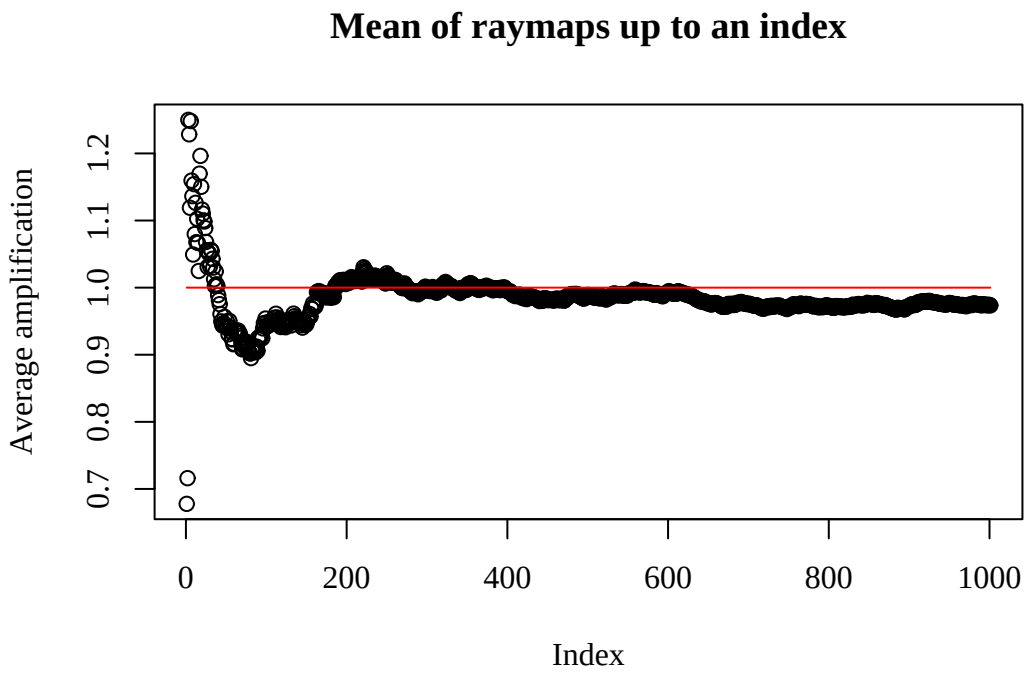

Figure 3.3.: *The average amplification of all raymaps up to an index. The average converges towards a value slightly lower than one (indicated by the red line). In this example, the pixel size is $P_d = 0.0025 R_E$ and the normalized surface mass density in stars is 0.5.*

3.2. Analytical functions

It is also possible to find an analytical function that represents a Distribution Function for the problem at hand.

3.2.1. Point Source

Refsdal (1970) discusses amplification of point sources in flat, static, but inhomogeneous universes. While the nomenclature is rather different from today's conventions, it is shown that for large amplifications, the probability of observing an amplification larger than some μ is proportional to μ^{-2} (Eq. (19b)). Except for a pronounced cutoff for small amplifications (deamplifications), the functions presented are smooth. Further, in this model (not in practice) point sources may be infinitely amplified.

In this section, I make use of the definitions and the discussion of distribution theory from Section 2.4.1 and make the rough approximation that the function found by

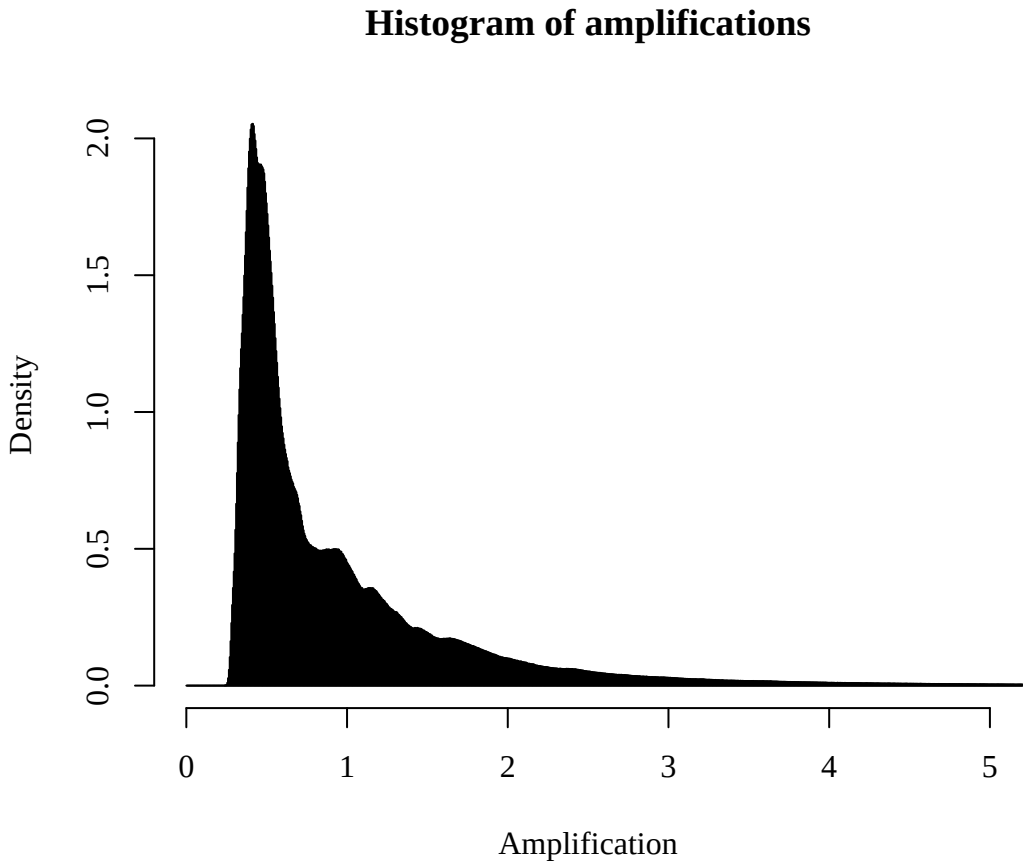

Figure 3.4.: *Histogram of amplifications. The pixel size is $P_d = 0.0025 R_E$. The figure only shows amplifications from 0 to 5; the density above 5 is very small. There are a total of 40 040 000 data points in this case, and the amplification bin width is 0.005.*

Refsdal (1970) holds down to a value μ_{min}, where I assume that the Probability Density Function has a sharp, lower cutoff, and the Distribution Function is zero.

I will find a Distribution Function which gives the probability of finding an amplification smaller than some μ. This implies that down to μ_{min}, the function must go as $A - k\mu^{-2}$ where A and k are constants to be determined.

The second condition cited in Section 2.4.1 is a normalization condition and implies that $A = 1$ since $k\mu^{-2} \to 0$ when $\mu \to \infty$. We have that

$$F(\mu) = \begin{cases} 0, & \text{for } \mu < \mu_{min} \\ 1 - k\mu^{-2}, & \text{for } \mu_{min} \leq \mu. \end{cases} \tag{3.1}$$

Since it is undesirable to have a finite probability accumulating for any points, we require that the Distribution Function is not only continuous from the right, but absolutely continuous, or

$$F\left(\mu \to \mu_{min}^{-}\right) = F\left(\mu \to \mu_{min}^{+}\right), \tag{3.2}$$

$$0 = 1 - k\mu_{min}^{-2}. \tag{3.3}$$

From this we find that

$$k = \mu_{min}^{2}. \tag{3.4}$$

The Distribution Function may now be written

$$F(\mu) = \begin{cases} 1 - \mu_{min}^{2}\mu^{-2}, & \text{for } \mu_{min} \leq \mu \\ 0, & \text{for } \mu < \mu_{min}, \end{cases} . \tag{3.5}$$

To find the corresponding Probability Density Function, we differentiate the above function and get

$$f(\mu) = \begin{cases} 2\mu_{min}^{2}\mu^{-3}, & \text{for } \mu_{min} \leq \mu \\ 0, & \text{for } \mu < \mu_{min}. \end{cases} \tag{3.6}$$

Because no lightray is lost or gained by lensing, the average amplification should be unity. Using this physically motivated condition, we find that we have a solution when $\mu_{min} = \frac{1}{2}$.

In conclusion, we have a Distribution Function for the amplification of a point source

$$F(\mu) = \begin{cases} 1 - \frac{1}{4}\mu^{-2}, & \text{for } \mu_{min} \leq \mu \\ 0, & \text{for } \mu < \mu_{min}, \end{cases} \tag{3.7}$$

and a corresponding Probability Density Function

$$f(\mu) = \begin{cases} \frac{1}{2}\mu^{-3}, & \text{for } \mu_{min} \leq \mu \\ 0, & \text{for } \mu < \mu_{min}. \end{cases} \tag{3.8}$$

I have implemented these functions in `R`, as Probability Density Function, Distribution Function and Quantile Function as well as a function that can be used to draw

random numbers from the distribution. Great caution is required when drawing numbers from a distribution such as this on a computer: Random numbers are generated by drawing numbers between 0 and 1 and evaluating the Quantile Function for the drawn number. However, as always on a computer, the numbers drawn between 0 and 1 have finite precision, and so there is not an infinite number of real numbers that can be drawn. For example, if the computer may not draw any numbers between 0.999999997 and 0.999999999, then there will be a quite large leap between the resulting numbers, in this case the former is 9128, the latter 15811. While no such amplifications will ever occur in nature, one should nevertheless take care so that this error does not influence the analysis done under the assumptions one makes. I have chosen the easiest fix: If the first number drawn is larger than 0.9999999, I draw again, this time with all representable numbers in the interval from 0.9999999 to 1. One could of course argue that numbers beyond 0.9999999 will seldomly be drawn since it is the tenth millionth quantile, but one should keep in mind that we are drawing amplification for millions of clouds, so we might see an effect of this in some line profiles.

3.2.2. Extended source

I will continue to use the Probability Density Function which falls off as μ^{-3}, but to take into account that an extended source will have a finite upper limit for the amplification, I will also introduce an upper cutoff. The distribution will have to satisfy this constraint as well as the constraints of the previous section.

We employ an analogous formalism to that above: A function

$$F(\mu) = A - k\mu^{-2} \tag{3.9}$$

is a Distribution Function between $\mu = \mu_{min}$ and $\mu = \mu_{max}$ if

$$0 = F(\mu_{min}) = A - k\mu_{min}^{-2}, \tag{3.10}$$

$$A = k\mu_{min}^{-2} \tag{3.11}$$

and

$$1 = F(\mu_{max}) = A - k\mu_{max}^{-2} = k\left(\frac{1}{\mu_{min}^2} - \frac{1}{\mu_{max}^2}\right). \tag{3.12}$$

Demanding an average amplification of unity implies

$$1 = 2k\left(\frac{1}{\mu_{min}} - \frac{1}{\mu_{max}}\right), \tag{3.13}$$

$$\frac{1}{k} = 2\left(\frac{1}{\mu_{min}} - \frac{1}{\mu_{max}}\right). \tag{3.14}$$

From Eq. (3.12) and Eq. (3.14) we see that

$$\frac{1}{\mu_{min}} + \frac{1}{\mu_{max}} = 2. \tag{3.15}$$

The Distribution Function becomes

$$F(\mu) = k\left(\frac{1}{\mu_{min}^2} - \frac{1}{\mu^2}\right), \tag{3.16}$$

$$= \frac{\frac{1}{\mu_{min}^2} - \frac{1}{\mu^2}}{\frac{2}{\mu_{min}} - \frac{2}{\mu_{max}}} \tag{3.17}$$

Only one of μ_{min} and μ_{max} may be a free parameter. Both have a relatively clear physical interpretation, so we may substitute either one in the above equation. It seems easier to substitute $\mu_{max} = \mu_{min}/(2\mu_{min} - 1)$ in the above equation, and with this, we see that

$$F(\mu) = \frac{\frac{1}{\mu_{min}^2} - \frac{1}{\mu^2}}{\frac{2}{\mu_{min}}(2 - 2\mu_{min})}, \tag{3.18}$$

$$= \frac{\frac{1}{\mu_{min}} - \frac{\mu_{min}}{\mu}}{4(1 - \mu_{min})}, \tag{3.19}$$

$$= \frac{\mu^2 - \mu_{min}^2}{4\mu^2\mu_{min}(1 - \mu_{min})}, \tag{3.20}$$

$$= \frac{1}{4\mu_{min}(1 - \mu_{min})} - \frac{\mu_{min}}{4(1 - \mu_{min})\mu^2}. \tag{3.21}$$

Some final rearrangement, and this may be formulated more precisely as

$$F(\mu) = \begin{cases} 0 & \text{for } \mu < \mu_{min}, \\ \frac{1}{4\mu_{min}(1-\mu_{min})}\left(1 - \frac{\mu_{min}^2}{\mu^2}\right) & \text{for } \mu_{min} \le \mu < \mu_{min}/(2\mu_{min} - 1), \\ 1 & \text{for } \mu \ge \mu_{min}/(2\mu_{min} - 1). \end{cases} \tag{3.22}$$

We note that this is only usable for $\frac{1}{2} \le \mu_{min} < 1$ and that it reduces to Eq. (3.7) in the case where $\mu_{min} \to \frac{1}{2}$.

Differentiating we get the corresponding Probability Density Function:

$$f(\mu) = \begin{cases} 0 & \text{for } \mu < \mu_{min} \text{ and } \mu \ge \mu_{min}/(2\mu_{min} - 1), \\ \frac{\mu_{min}}{2(1-\mu_{min})\mu^3} & \text{for } \mu_{min} \le \mu < \mu_{min}/(2\mu_{min} - 1). \end{cases} \tag{3.23}$$

These functions are then implemented in `R` along with their corresponding Quantile Function. A plot of the Distribution Function is shown in Fig. 3.5. To be able to use either μ_{max} or μ_{min} as a free parameter, the `R`-function also incorporates Eq. (3.15), and assumes that the input parameter is μ_{max} if greater than unity and μ_{min} if less than unity. To be on the safe side, it also takes advantage of the algorithm I developed to ensure that no great jump occurs because of the discrete representation of numbers, as discussed above.

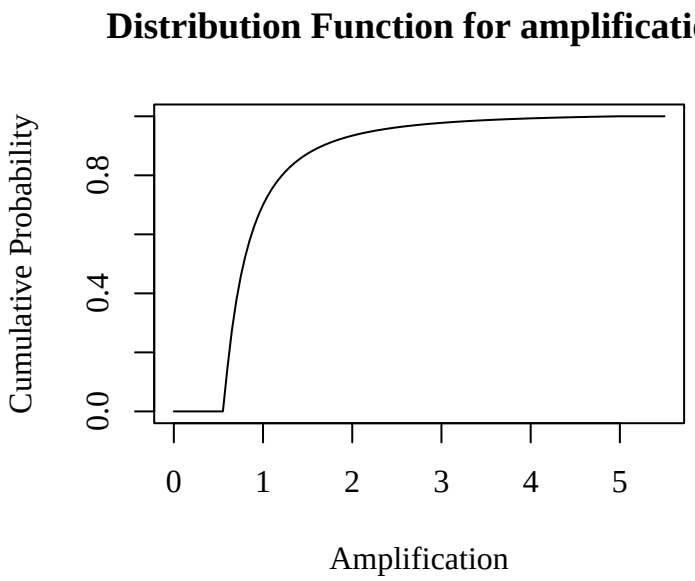

Figure 3.5.: *The Distribution Function for amplification using Eq. (3.22) and a maximum amplification of 5.*

Finally, when these analytical distributions are used in the sequel, I will mostly be using μ_{max} as the free parameter. Therefore, it falls naturally to identify these analytical distributions by citing their maximum amplification. I will do this by using the symbol $\mu_{max,an}$. If the distribution for a point source as found in Section 3.2.1 is used, it is identified as $\mu_{max,an} = \infty$.

3.3. Comparing numerical and analytical distributions

To compare the numerical distributions with the analytical distributions for the extended source, I have made a histogram with a detail from Fig. 3.2, and superimposed a histogram with randomly generated numbers from the distribution in Section 3.2.2, using the maximum amplification in Fig. 3.2, namely 26.6 as input parameter, and the same total number of data points. It is the behavior for large amplifications that is the most interesting, so I have plotted amplifications from 1 to 6, as shown in Fig. 3.6.

We see that the histograms agree well in this region. From Eq. (3.23) we know that the black histogram falls as μ^{-3}, and this implies that the numerical distribution found by using `cmap` agrees well with the result of Refsdal (1970) that was the core assumption of Section 3.2.

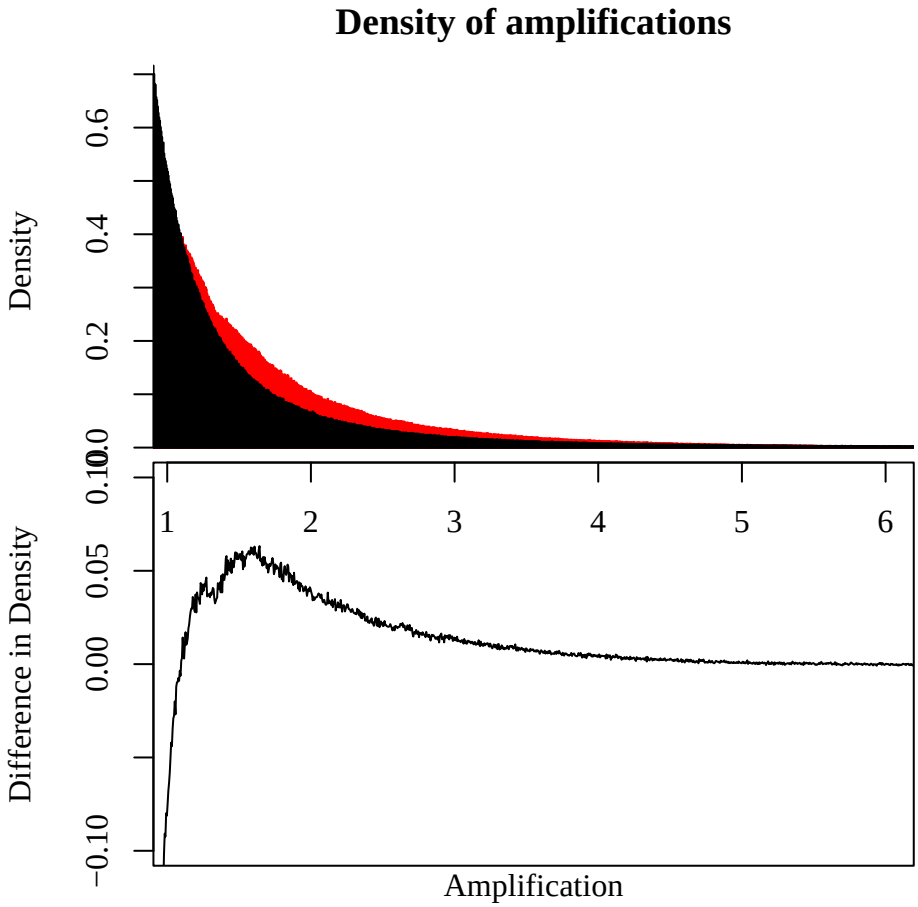

Figure 3.6.: *The upper panel shows two histograms of the amplification from 1 to 6. The red histogram is the numerical distribution for* $\kappa = 0.5$*, the black histogram is for the analytical distribution found in Section 3.2.2. The lower panel shows the difference between the two densities. There are a total of 9 960 000 data points for each set and the amplification bin width is 0.005.*

4. Basic Spectral Line Profiles

4.1. Assumptions

There is a large number of physically motivated assumptions that have been made and some issues that are a matter of personal preference that needs elaboration.

4.1.1. Clouds are physically identical

First, I assume that the clouds are physically identical. By this, I mean that they are all of the same size, have the same temperature, have the same optical thickness, same element abundances, etc. Consequently, the intensity of a cloud will be the same for all clouds, and given a cloud's projected position in a caustic network any cloud at the same projected position would be amplified by the same amount.

4.1.2. Emitted light is monochromatic

I will study one emission line at a time and within a line I assume that the light emitted is monochromatic, that is, light from a single cloud contributes to only one bin. The importance of this assumption lies in that I may associate a cloud with a given velocity with a specific bin, and then the rough appearance of the spectral line profile is that of the distribution of cloud velocities. In the following I will say that "a cloud is in a bin" when a cloud has such a velocity that its radiation will hit a specific bin.

4.1.3. Cloud velocity model

The distribution of cloud velocities gives the first order line profile, in the sense that it is the Doppler shift of clouds in motion that gives the line's appearance, but there are many other mechanisms that alter this somewhat to give the final observed line profile, such as photon noise, lensing, etc. While the line profile depends strongly on the kinematics of the BLR, I choose to simplify the analysis by ignoring this and use a Gaussian function as a zeroth order line profile. Krolik (1999) writes that a typical BLR-line resembles a linear combination of two Gaussians, a single Gaussian serves as a special case of this.

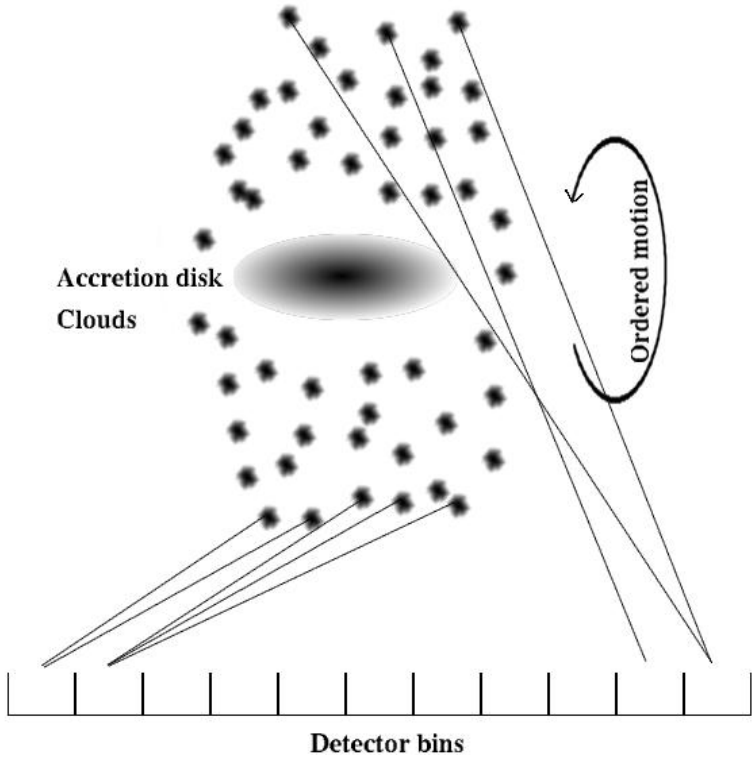

Figure 4.1.: *Clouds around an accretion disk emitting photons that are detected in different bins. The clouds are in some purely hypothetical ordered motion around the quasar core, and with the shown motion, photons coming from the clouds at the bottom of the figure are blueshifted and contribute photons to the bins in the left of the figure while photons coming from clouds at the top are redshifted and so contribute to the bins at the right hand side, as depicted by the lines. Further, clouds closer to the axis of rotation contribute to bins closer to the center of the detector.*

In Fig. 4.1 the clouds are in ordered motion around the quasar core, specifically they are in a polar orbit, perpendicular to the accretion disk. This is entirely hypothetical and not even plausible, and is used merely as an illustration. The main point is to illustrate how a line is broadened by cloud velocities.

To reflect the fact that there will be fluctuations in the number of clouds in a given bin, this number is drawn from the Poisson distribution with a parameter that equals the fraction of the total number of clouds as given by the Gaussian function. The resulting profile is illustrated in Fig. 4.2, upper left panel.

4.1.4. Clouds are lensed independently

I assume that clouds are lensed independently, in the sense that the amplification of one cloud is uncorrelated with the amplification of the other clouds. In a model where a realistic kinematical model had given the first order line profile, as opposed to just using a Gaussian as mentioned in Section 4.1.3, this would have been an unacceptable assumption, since in many models, the velocity of a cloud would be closely related to its position around the quasar core as illustrated in Fig. 4.1, and so emission from neighboring clouds may end up in the same or neighboring bins. Since neighboring clouds would also be close when projected into a caustic network, they would have strongly correlated amplifications. However, since we do not assume a kinematical model, it follows naturally to assume that clouds are lensed independently.

4.2. Flux density from lensed clouds

The amplification is the ratio of flux densities between a lensed and an unlensed image. This was used in Eq. (2.9) to show that the amplification depends only on the solid angle subtended by lensed and unlensed images of the source. Since we are discussing the narrow bands of a spectrum, we naturally find it most convenient to use the *specific flux density*, i.e. the flux density per unit wavelength (or frequency). I use a subscript ν to indicate specific flux density.

Since clouds are physically identical (Section 4.1.1), the specific flux density of radiation emanating from a cloud, $\mathcal{F}_\nu$, is the same for all clouds. The specific flux density of radiation impinging on the detector from a cloud is then

$$f_\nu = \mu \mathcal{F}_\nu. \tag{4.1}$$

Let us say that we have a telescope with a primary mirror with an area of A, and we take an exposure of duration τ. The energy of the radiation then impinging on the detector during such an exposure, the specific flux density per photon energy from a single cloud, can be written as

$$\varphi_\nu = \frac{f_\nu \tau \lambda A}{hc}, \tag{4.2}$$

where λ is the wavelength of the radiation, h is Planck's constant and c is the speed of light. Since the lines cover a relatively short wavelength interval, I will for simplicity use a fixed value for the wavelength.

While the physical identity of clouds dictate that the specific flux density is the same for all clouds, the number of photons *detected* from clouds are not. I stated on page 26 that detection of photons is a counting process and it is therefore conventional to assume that the number of photons detected is a Poisson-distributed variable with a parameter that is simply the specific flux density per photon energy multiplied by the sensitivity S: $\varphi_\nu S$.

Now, this parameter has another obvious interpretation: It is the average number of photons detected from a cloud. It is of course possible to think in terms of photons all the way from emission to detection, but one must then take care to avoid some pitfalls along the way. One such pitfall is to draw a number of photons from the Poisson distribution and multiply it with some amplification factor. This will lead to absurdities, such as for example amplifying 3 photons by a factor 1.2, getting 3.6 photons.

It is important to realize that amplification arises because light rays are collected from a larger solid angle than it would if unlensed. In statistical terms, this means that the "success rate" is changed, and consequently, that the Poisson parameter is changed. This is exactly the same result as one gets by the specific flux density-reasoning above.

Having been in the stated pitfall, I feel more confident when insisting on using specific flux density all the way to the detector, using the concept of photons only in the detection process itself.

4.3. Flux density from continuum

I assume that the continuum is unlensed or lensed independently from the clouds, and that the detected radiation from the continuum is also Poisson distributed. It is reasonable to assume that the continuum can be lensed independently from the clouds in light of the discussion in Section 4.1.4 and that the clouds are spatially separated from the continuum source.

4.4. Free parameters

In generating line profiles, there are naturally a number of parameters that are reasonable to vary, and a number of parameters that can be assumed to be fixed at some value, and finally, the parameter space has some degeneration, so that we may vary one parameter and fix others without loss of generality. The challenge is to determine which parameters belong to each group, and set the parameters that may be fixed to reasonable values. It is also wise to attempt to keep the number of varied parameters to a minimum, at least for the initial analysis.

Let us first discuss the following parameters: detecting instrument sensitivity, aperture of telescope, exposure time and total flux density of the quasar. All these parameters are degenerate; we can always compensate for a weak quasar by taking longer exposures, using a more sensitive instrument or a bigger telescope.

I have in fact fixed all these parameters, and instead I focus on the question "How many counts can I at most reasonably expect to get in the line center?" The count in the line center, N_{max}, will be, or will be very close to the overall maximum count for a line profile, and is therefore a parameter I can adjust according to my expectations in an observational campaign. Yet, it is also a parameter that I fix.

Another important parameter which I fix is the width of the bins in nanometers, denoted W_b.

The parameters I keep entirely free are the Equivalent Width of the emission line (EW), the total number of clouds (N_c) and the number of bins that the line covers (N_b). Additionally, I will deal with different lens models, which gives different distributions for the amplification, as discussed in Chapter 3.

Using the above mentioned parameters, I may compute the quantities that are necessary for generating the line profiles.

First, we need to find the Poisson parameters as given by the flux from a cloud or the continuum, divided by the energy of a photon. The parameter for the continuum is given analogous to Eq. (4.2), but first I find the specific flux density per photon energy:

$$\Phi_\nu = \frac{F_\nu \tau \lambda A}{hc}, \tag{4.3}$$

where F_ν is the impinging specific flux density from the continuum.

Since the specific flux density per photon energy from a cloud, denoted φ_ν, and the specific flux density per photon energy from the continuum is related through the Equivalent Width by

$$EW = \frac{\varphi_\nu W_b N_c}{\Phi_\nu}, \tag{4.4}$$

and since most of the light comes from the continuum anyway, I will use the specific flux density from the continuum as a measure of quasar brightness which is a nice fixed input parameter.

4.5. Generating line profiles

With the specific flux density per photon energy known for both the continuum using Eq. (4.3) and a cloud by solving Eq. (4.4) for φ_ν, we may now describe the generation of line profiles. This involves choosing appropriate values for parameters and drawing numbers from the appropriate distributions of the random variables.

Generation of line profiles has been implemented in `R`. The code is included in Appendix A. The power of `R` makes the code very simple, generating line profiles requires effectively around 20 lines of code.

Initially, I create the zeroth-order line profile, as I started describing in Section 4.1.3: First, I take a normalized Gaussian with center at zero and a FWHM of σ_l (in units of bins). The Gaussian is divided into N_b bins and then multiplied with the total number of clouds N_c. We now have a smooth function where the total number of clouds is distributed (non-randomly) in the bins according to the Gaussian. To reflect the fact that whether a cloud ends up in one bin or the other is a stochastic process, I draw a number from the Poisson distribution with a parameter given by the Gaussian. I take this to be the number of clouds in a bin. The upper left panel of Fig. 4.2 shows the

resulting first-order profile after this step. This works well as long as the number of bins and clouds are reasonably large. To see why, imagine for example two bins and one hundred clouds; for each bin we draw a number from a Poisson distribution with a parameter of e.g. 50. It is not unlikely that both drawings yield a higher number than 50, and so we have gotten more clouds than we started out with. Such errors cancel out when dealing with many bins. To correct for the problem, one could use a binomial (or multinomial) distribution. Implementing this in R is very easy but the computing time needed to draw a number from the binomial is typically three times of that needed to draw from the Poisson, with little won in accuracy of the analysis in the cases of our interest.

Secondly, it is convenient to calculate the exposure time from the chosen threshold of counts. Since this only needs to be approximate, I take the exposure time to be

$$\tau = \frac{N_{max}}{S(M\varphi_\nu + \Phi_\nu)}, \tag{4.5}$$

where M is the maximum number of clouds in the profile.

I concentrate on a case where we have four quasar images, and obtain a spectrum of one emission line in each. Further, the four emission lines have the same Equivalent Width, and the number of clouds in a given bin is the same for the four spectra. The latter is expected to be the case in objects where the time it takes for a cloud to change the velocity so much that its photons hit a different bin is shorter than the time delay between images. In this case, I use the same first-order line profile, i.e. the same Gaussian with Poisson "noise" representing the number of clouds in a bin, for the four line profiles. The influence of using lines with the same Equivalent Width is that a straightforward contingency table (see Section 2.4.3) can be used, but unfortunately, the results from this analysis can only be compared to observations in the rare case when the four lines do indeed have similar Equivalent Width. It is nevertheless an interesting case.

Now, I create a matrix where the rows are the four identical line profiles (a column represents one particular bin), and for each element in the matrix, I do the following:

1. I create an array of length corresponding to the number of clouds in that bin. For each element, I draw the amplification from a selected amplification distribution, multiply it with the specific flux density per photon energy from an unlensed cloud[1] so the elements in this array represents the specific flux density per photon energy φ_ν, from each cloud impinging on the detector. The lower left panel of Fig. 4.2 shows how the line appears if I sum up the elements of this array for every bin for one line profile.

2. To find how many photons is detected in a bin, I use the above array, and for each element, I multiply it with the sensitivity S to reflect that only a fraction of

[1] To preserve generality in the R-code, this is not exactly how it is programmed, but the difference is of no importance.

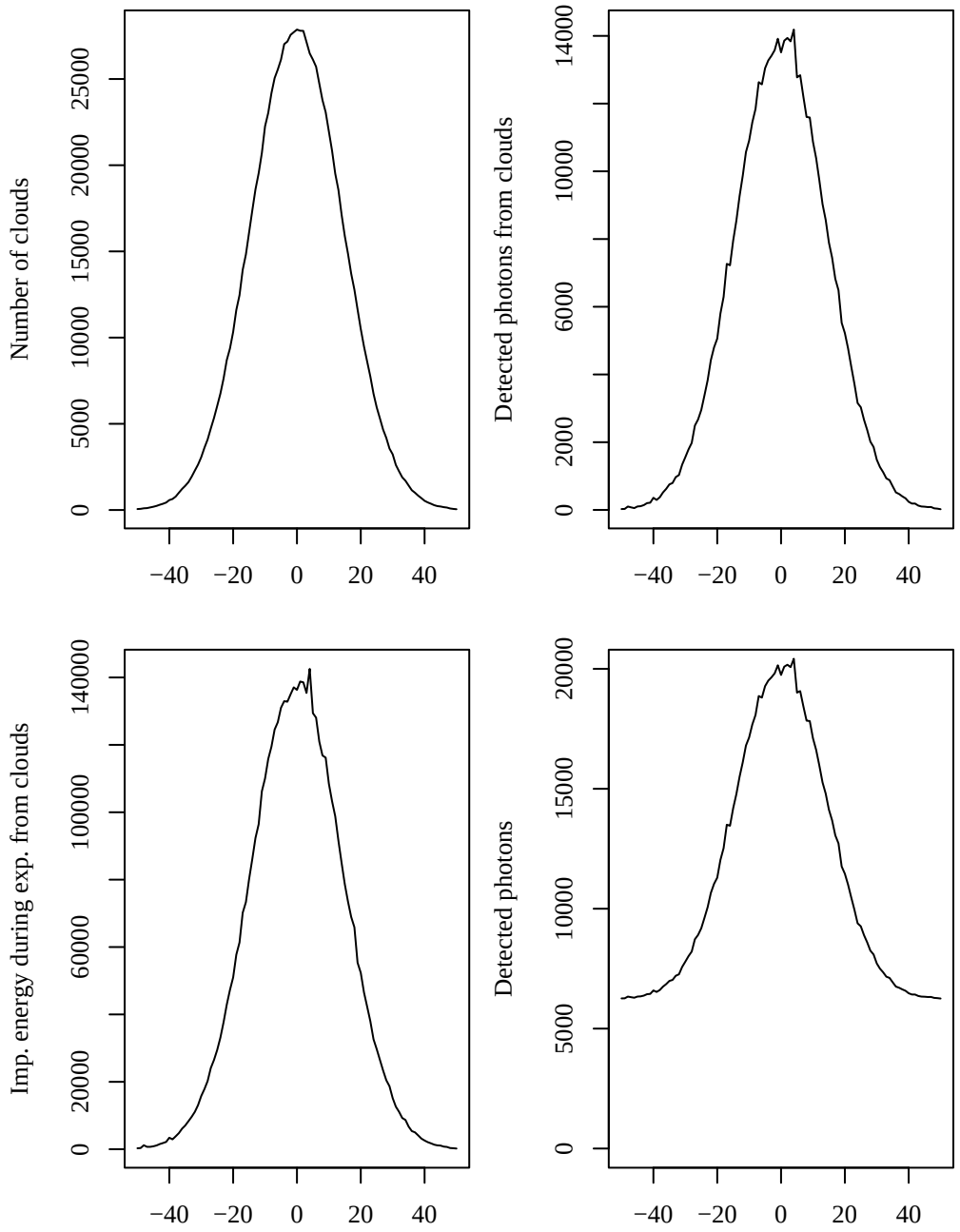

Figure 4.2.: *The generation of a line profile. The graphs are produced assuming* $EW = 8$ *nm,* $F_\nu = 3 \times 10^{-18} Wm^{-2}Hz^{-1}$, $N_{max} = 20000$, $W_b = 0.1$ *nm,* $A = 8$ *m,* $N_b = 100$, $S = 0.1$ *and* $N_c = 10^6$. *The upper left panel shows the number of clouds in each bin, having the shape of a Gaussian with Poisson "noise" added. The lower left panel shows the energy from clouds impinging on each bin during an exposure in units of photon energy. The upper right panel shows the photons detected from clouds only, and in the lower right panel, the photons from the continuum has been added.*

the impinging specific flux density will be detected, and draw a number from the Poisson distribution, with $\varphi_\nu S$ as its parameter.

3. Then, I sum all the elements of the array; this number is the total number of photons from all the clouds in that bin. See the upper right panel in Fig. 4.2 for an illustration.

4. Finally, I add another number drawn from the Poisson distribution with $S\Phi_\nu$ as parameter, which is the number of photons in the bin from the continuum. We have finally got complete line profiles, and one is depicted in the lower right panel of Fig. 4.2.

When this has been done for all the elements in the matrix, we have a new matrix with four different line profiles. The line profiles represent the number of photons in each bin, for each of the four quasar images.

4.6. Input values for the parameters

Many different simulations will be conducted, with different input parameters. I have, however, chosen a "core set" of simulations. This core set consists of simulations for several combinations of different values for the Equivalent Width, the total number of clouds and the number of bins as well as using a selection of different amplification distributions.

Table 4.1.: *Parameter values as used in the core set of simulations.*

Description	Symbol	Value(s)
Equivalent Width	EW	8 – 25 nm
Total number of clouds	N_c	$10^6 - 3 \times 10^7$
Number of bins	N_b	100 – 1000
Specific flux density from continuum	F_ν	3×10^{-18}W m^{-2} Hz^{-1}
Maximal number of counts	N_{max}	20000
Width of detector bins	W_b	0.1 nm
Aperture of telescope	A	8 m
Sensitivity of detecting instrument	S	0.1
Wavelength of radiation	λ	550 nm

In Table 4.1, the free parameters that are being varied are tabulated, along with the parameters that are fixed. A short description is given, and the symbol used is tabulated. For the free parameters, the interval used is indicated, while for the fixed parameters, the value is given.

There is one quirk in the code: To get symmetric line profiles, I allow the code to add a bin to the given value if the value is an even number. Thus, the input value may be e.g. 1000 but the number of bins used is 1001.

4.7. Relation to previous studies, revisited

We are now better prepared to understand the difference between the present study and the studies done by Nemiroff (1988) and Schneider & Wambsganss (1990).

First, both studies take a BLR with internal structure and convolve the sources with a raymap. In the first paper, a Chang-Refsdal lens is used, in the latter, realistic microlens models with many point masses are used.

Both studies conclude that in the case where clouds are in purely random motion, it will be hard to detect lensing at all. It is the ordered motion assumed by the different models investigated that makes the rather dramatic effects possible. These dramatic effects are best exemplified in Fig. 4 of Schneider & Wambsganss (1990). Changes in the line profile of this magnitude are very unlikely to occur with the assumptions I make.

It is clear that the effect of random cloud motions is preventing the amplification of the clouds in a bin from being accumulated. That is the very same effect of the assumption that clouds are lensed independently. Two clouds that are close together in space, where they may have strongly correlated amplifications, may have very different velocities, and so they will end up in different bins, and therefore, the effect of lensing in a bin will not add up.

It is clear that the assumption of independently lensed clouds is a very rough approximation, and it is not very likely to be correct. While this assumption, combined with the use of a simple Gaussian function makes the forthcoming analysis possible, it can also be looked upon as a worst-case-scenario in this context: If it is true, the effects of microlensing are very small compared to those reported by e.g. Schneider & Wambsganss (1990). Therefore, if lensing can be detected in this regime, it is quite likely to be detectable in regimes where clouds are not lensed independently and where there is some ordered motion. It is, of course, possible to imagine situations where the detection of microlensing of clouds is even more difficult. Actually, I find one such situation in Section 7.3 discuss it in Section 8.7 and another possible situation in Section 7.5. When compared to previous studies the general model that is being studied in this thesis is found to be analogous to what has previously been considered the worst case.

Finally, the strength of the present analysis will lie in its statistics — that I am able to simulate so many line profiles and make so many tests that a good view of the distributions of the underlying random variables can be achieved. This would probably have been too expensive to do using the approach of previous papers.

5. Description of the test

The contingency table and its test statistic were explained in general terms in Section 2.4.3. In this chapter, I will describe the use of contingency tables for our particular problem. We will examine the test's ability to identify bins with highly amplified clouds. Also, the rejection threshold will receive some attention, and we will evaluate the test according to the principles set forth on page 29, and note some conditions where the test performs well, and some conditions where it does not.

First, we must determine the exact nature of the problem we wish to solve: There are two distinctly different questions that may be asked: "Given an observation, what conclusion should be drawn?" and "Under what conditions may lensing be detected reliably?" Either question must be answered by using a statistical test, and the same test may be employed to answer either question. However, the former question represents an *a posteriori* problem, the latter is an *a priori* problem.

No actual data has been obtained for this thesis, and therefore, it is natural that I will focus on the *a priori* problem of determining under what conditions we may detect microlensing.

5.1. Applying the contingency table

It is time to apply the theory of Section 2.4.3 to the specific line profiles. Section 4.5 was concluded with stating that a matrix containing four line profiles had been generated by following the procedure described in that section. Let us assume that we now have this matrix. In Eq. (2.48), a matrix is given with O_{ij} as the observed number of counts in a bin for four line profiles, and consequently we now identify this matrix as the matrix we got by following the procedure of Section 4.5: O_{ij} contains the four line profiles, one row for each.

I have made an example contingency table in Table 5.1 with four line profiles and with 9 bins. The marginal sums are also included. The row sums represent the relative brightness of the images, and the column sums, when divided by the number of images, can be thought of as an averaged spectrum.

Next, the expected number of counts would have to be estimated, according to Eq. (2.49). In our case, the number of rows are $k = 4$ and the number of columns

Table 5.1.: *Example contingency table: Observations matrix O_{ij} containing photon counts. The rows can be thought of as spectra for each image and the columns are bins.*

	1	2	3	4	5	6	7	8	9	Sum
1	278	855	4601	13840	19909	13767	4469	836	269	58824
2	252	863	4487	13594	19994	13661	4542	809	264	58466
3	239	851	4437	13595	19986	13792	4532	821	257	58510
4	264	881	4579	13494	20010	13634	4384	877	266	58389
Sum	1033	3450	18104	54523	79899	54854	17927	3343	1056	234189

are $l = N_b$, so Eq. (2.49) becomes

$$\hat{E}_{ij} = \frac{\left(\sum_i^4 O_{ij}\right) \cdot \left(\sum_j^{N_b} O_{ij}\right)}{\sum_{\text{all}ij} O_{ij}}. \tag{5.1}$$

Using this and the numbers from Table 5.1, we may compute a table that looks like Table 5.2.

Table 5.2.: *Example contingency table: Expected photon counts $\hat{E}_{ij}$.*

	1	2	3	4	5	6	7	8	9
1	259	867	4547	13695	20069	13778	4503	840	265
2	258	861	4520	13612	19947	13694	4476	835	264
3	258	862	4523	13622	19962	13705	4479	835	264
4	258	860	4514	13594	19921	13676	4470	833	263

The expected number of counts is easiest to understand in the context of the averaged spectrum and the relative brightness. The averaged spectrum is the best estimate of the unlensed profile from the data as the effects of lensing on the different images are expected to be uncorrelated and average out. Therefore, the expected number of counts is the averaged spectrum, corrected so that the relative brightness is preserved (though in our case, they should be almost identical).

Further, we noted in Eq. (2.50) that H_0 is rejected when the χ^2-test statistics exceeds the rejection threshold. Hereafter I shall use T to denote the rejection threshold, and we may now write Eq. (2.50) as

$$\Theta > \chi^2_{3(N_b-1),1-\alpha} = T. \tag{5.2}$$

In our case, it might be argued that we *do* know exactly what to expect for the null hypothesis, namely the photon counts in each bin with no lensing, and that we could simply use that instead of estimating the expected counts. However, in any realistic situation, we cannot simulate the line profile, so the expected counts would have to be estimated, and estimating the counts in this case is necessary for the formalism to be applicable.

Further, in any realistic situation one can only know the unlensed profile from an unlensed image of the actual object, which is rarely available and cannot be determined *a priori* to be unlensed. Therefore, when dealing with real data one has to estimate the unlensed profile and it would be nonsensical to do otherwise with the simulated data.

Table 5.3.: *Example contingency table:* Θ_{ij}.

	1	2	3	4	5	6	7	8	9	Sum
1	1.323	0.155	0.632	1.531	1.278	0.009	0.256	0.016	0.053	5.25
2	0.135	0.003	0.237	0.023	0.111	0.082	0.987	0.785	0.001	2.36
3	1.411	0.139	1.640	0.054	0.029	0.555	0.630	0.242	0.177	4.88
4	0.161	0.504	0.943	0.734	0.400	0.132	1.641	2.271	0.028	6.81
Sum	3.030	0.802	3.450	2.340	1.820	0.778	3.510	3.310	0.258	19.3

The Θ_{ij}'s are found using Eq. (2.48) and the χ^2-test statistic Θ is computed by summing over all elements, as described in Section 2.4.3. An example where Θ_{ij} is computed from the data in Table 5.1 is included in Table 5.3, and we note that Θ is in fact the figure in the last cell of the table. The next section will outline how this is done in practice.

5.2. Implementation

Both generation of line profiles and the χ^2-test has been implemented in `R`, the system mentioned in Section 2.4.5. The contingency table based χ^2-test is a conventional test, and is distributed with the `R` system in the `ctest`-library ("classical tests"), which is well documented in the publicly available `R` manuals. Both the `R` χ^2-test and my own code are programmed as generalized functions that can be called with different parameters, and so the code becomes very transparent. My code is included in Appendix A.

With the line profiles generated, running the test is done by a single line call. For example, testing whether lensing can be detected in a situation with $N_c = 5$ million clouds, an Equivalent Width of $EW = 18$ nm, $N_b = 500$ bins and $\mu_{max,an} = \infty$ (the amplification distribution for a point source) can be done by invoking the command:

```
chisq.test( lineprofiles (5000000, 18, 500, structure(0, class="point")))
```

In this example, R responds:

```
        Pearson's Chi-squared test

data:  lineprofiles (5e+06, 18, 500, structure(0, class = "point"))
X-squared = 1918.913, df = 1500, p-value = 9.235e-13
```

Interpreting the output is straightforward: 1918.913 is the total χ^2-test statistic Θ, **df** is the degrees of freedom, given in Section 2.4.3 as $(k-1)(l-1)$ or in our case $3(N_b-1)$. Since N_b is increased by one by the code to get a symmetric profile, the degrees of freedom in this example is indeed 1500. The p-value depends monotonically on the test statistic for any given degrees of freedom, so I will cite only the test statistic unless I need to compare tests done with different N_b.

Furthermore, if one wishes to run multiple tests, either for a single set of parameters or for a parameter space, also at most a couple of lines are needed. Results may also be assigned to objects (variables), allowing storing and easy access for later analysis, and such an object will also contain more information than the above report. This is convenient to note, because a hypothesis test object will contain, among other things, the line profiles themselves. These will be in the form of the observation matrix O_{ij}, and the expected counts $\hat{E}_{ij}$, calculated by the R system itself using an implementation of Eq. (5.1). The Θ_{ij}'s are not saved, but as we shall use them in the next session, I shall show how they can be obtained as an example of the matrix calculation capabilities in R:

```
statistics <- (test$observed - test$expected)^2 / test$expected
```

where statistics is Θ_{ij}, test$observed is O_{ij} and test$expected is $\hat{E}_{ij}$.

5.3. Identifying source of variation

While most of the attention in this study is devoted to finding out under what conditions the variability caused by lensing is large enough to be distinguished from the variability in the unlensed case for the line profile as a whole, it is also interesting to investigate where in the line profiles the greatest deviation arises.

To study this, one might compare the amplifications of the clouds in each bin in each of the line profiles with the corresponding Θ_{ij}'s. We remember that the Θ_{ij}'s represent the contribution from each bin and image to the total test statistic. As this is impractical due to the large amount of data that would be produced by such an approach, I have recorded some characteristics. The first characteristic is the maximum amplification of cloud k for profile i and bin j. If μ_{ijk} is used for the amplification for each cloud in each bin and for the four profiles, then I shall use

$$\mu_{max,ij} = \max_k \mu_{ijk} \tag{5.3}$$

to denote the maximum amplification recorded among the clouds in profile i and bin j.

The second characteristic I record is the absolute value of the mean of the amplifications subtracted by 1 for each bin and profile, or

$$\mu^a_{ij} = \left|\left\langle \mu_{ijk} - 1 \right\rangle_k\right|. \tag{5.4}$$

That is, I average over all clouds for each bin and profile, and I shall use μ^a_{ij} to signify this characteristic in the following.

The latter is designed to be a measure of how much the amplification deviates from the expected amplification (i.e. 1, no amplification) in a bin, since amplification will contribute positively to this number while a deamplification will contribute negatively. Since Θ_{ij} will have a positive peak for both exceptional dips and peaks in the spectra, the absolute value of the mean is then used.

It is important to observe that in the following subsections I will use both the test statistics for the bins, Θ_{ij} (for a given but unimportant value of i) and the test statistic for the test as a whole Θ (which is the sum the test statistics for each bin and image).

5.3.1. Point source

Let us first, for illustrative purposes, consider the unrealistic case where the emitting source is a point source, that is, we allow the amplification to be infinite. We make use of the amplification distribution developed in Section 3.2.1, in the following identified by $\mu_{max,an} = \infty$.

In Fig. 5.1, a line profile is plotted in black as well as a graph showing the maximum amplification $\mu_{max,ij}$ in each bin. An extreme peak in bin 10 is easily seen, and it is clearly associated with the high amplification in this bin, as shown by the red line. Our primary concern now is to demonstrate that the test is capable of identifying lensing, and even identifying in what bin(s) this lensing occurs. The blue line is Θ_{ij} and we see that there is excellent agreement between high values for Θ_{ij} and the maximum amplifications, and indeed the correlation coefficient for $\mu_{max,ij}$ and the Θ_{ij} is 0.93.

Since I am currently testing whether four spectra obtained from four quasar images can be told apart, it is also interesting to examine the correlation between Θ_{ij} and $\mu_{max,ij}$ and μ^a_{ij} along with the χ^2-test statistic Θ for the entire test (the test of whether the hypothesis that no lensing can be seen, should be rejected). I have done many such tests for the same parameters as above and tabulated the results in Table 5.4. The graphs in Fig. 5.1 are taken from a line profile in test number 5 in this table.

As can be seen, the correlation coefficients for Θ_{ij} vs. $\mu_{max,ij}$ are mostly quite large, which means that in this case, the test is excellent in identifying bins with high amplification. The correlation coefficients for μ^a_{ij} are very variable but may be acceptable in some cases. In Table 5.5 I have tabulated the correlation coefficients for the columns of the above table, so we can study the covariation of the variables in Table 5.4.

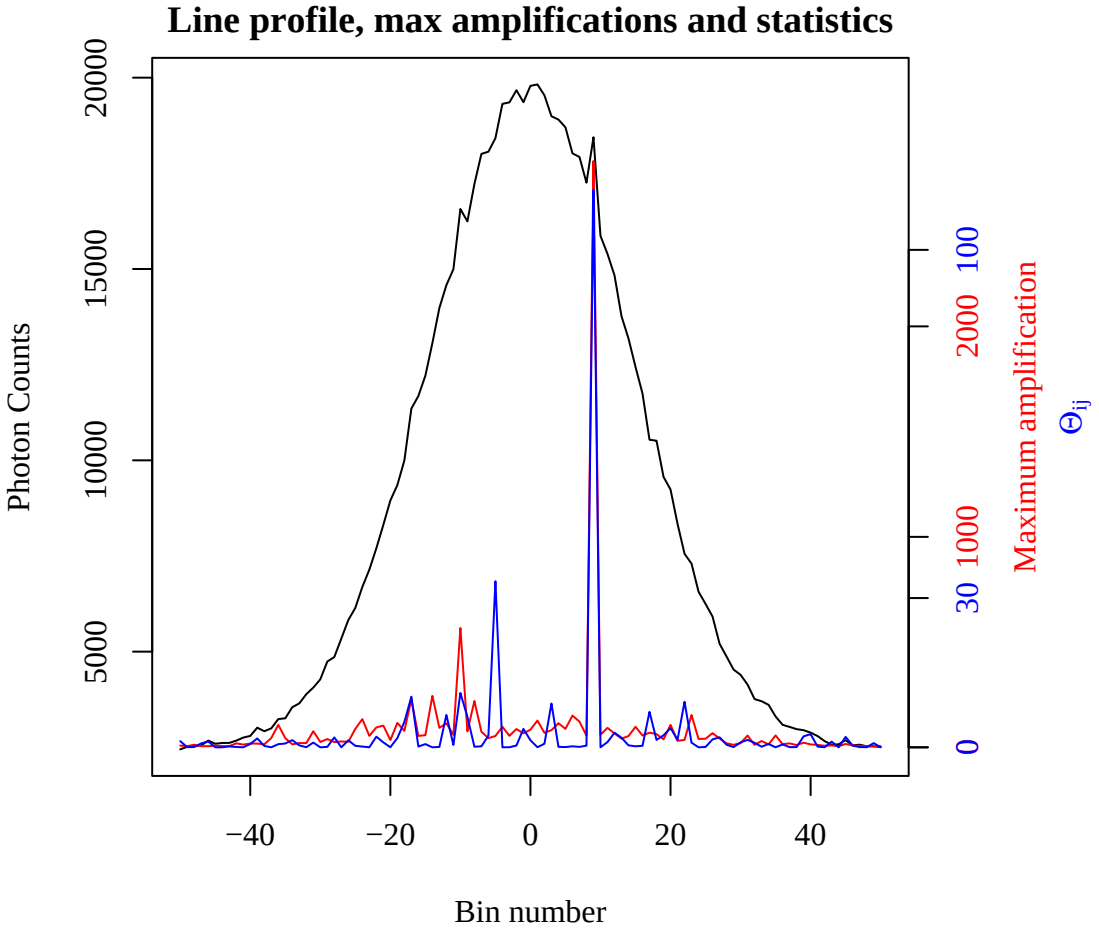

Figure 5.1.: *Photon counts, the highest amplification $\mu_{max,ij}$ and Θ_{ij} for each bin. The line profile in black is produced assuming 1 million clouds, that the line has an Equivalent Width of 25 nm, and covers 101 bins. The $\mu_{max,an} = \infty$ amplification distribution is used. The maximum amplification in each bin is in red, and Θ_{ij} for that bin is in blue.*

5.3.2. Numerical amplification distribution

In this section I will repeat the above analysis using the numerical amplification distribution that is based on raymaps with pixel size of $P_d = 0.0025 R_E$ described in Section 3.1.2.

Naturally, we cannot expect to get as dramatic effects in the case of an extended source. First, the maximum amplification is the amplification of a single cloud, and with millions of clouds and finite amplification, any bin is not likely to be dominated by a single cloud. Second, an intrinsic weakness of using the maximum amplification is that the more clouds there are in a bin, the higher will the mean of the maximum

Table 5.4.: *Correlation between Θ_{ij} image and $\mu_{max,ij}$ and μ^a_{ij} for $\mu_{max,an} = \infty$. The total χ^2-test statistics Θ for a test is also tabulated.*

Test number	Θ	Correlation Coefficient	
		$\mu_{max,ij}$	μ^a_{ij}
1	918	0.74	0.94
2	794	0.85	0.22
3	788	0.77	0.49
4	672	0.58	0.18
5	1052	0.91	0.20
6	735	0.76	0.14
7	618	0.77	0.21
8	2704	0.96	0.45
9	502	0.19	−0.03
10	722	0.48	0.29

Table 5.5.: *Correlation between Θ, the correlation coefficients for $\mu_{max,ij}$ and correlation coefficients for μ^a_{ij}, as tabulated in Table 5.4.*

	$\mu_{max,ij}$	μ^a_{ij}
Θ	0.55	0.31
$\mu_{max,ij}$	1.00	0.40

amplification become. Therefore, the absence of high correlation coefficients need not be an indication of shortcomings of the test, but rather an indication of the shortcomings of the characteristics I have recorded about the amplification in each bin.

Another weakness of relying on the maximum amplification to correlate with Θ_{ij} is that the Θ_{ij} identifies not only bins that have exceptionally *many* photons, but also bins that have exceptionally *few* photons. In such cases, several clouds may have been deamplified.

After some preliminary analysis showed that the correlation coefficients of Θ_{ij} and μ^a_{ij} was close to zero, I tried to address the above mentioned problem by replacing it with recording the maximum absolute value of the base-10 logarithm of the amplification, that is

$$\mu^l_{ij} = \max_k \left|\log_{10} \mu_{ijk}\right|. \tag{5.5}$$

This too has its weaknesses, but as we shall see, it has its virtues as well. I shall use μ^l_{ij} to denote this characteristic in the following.

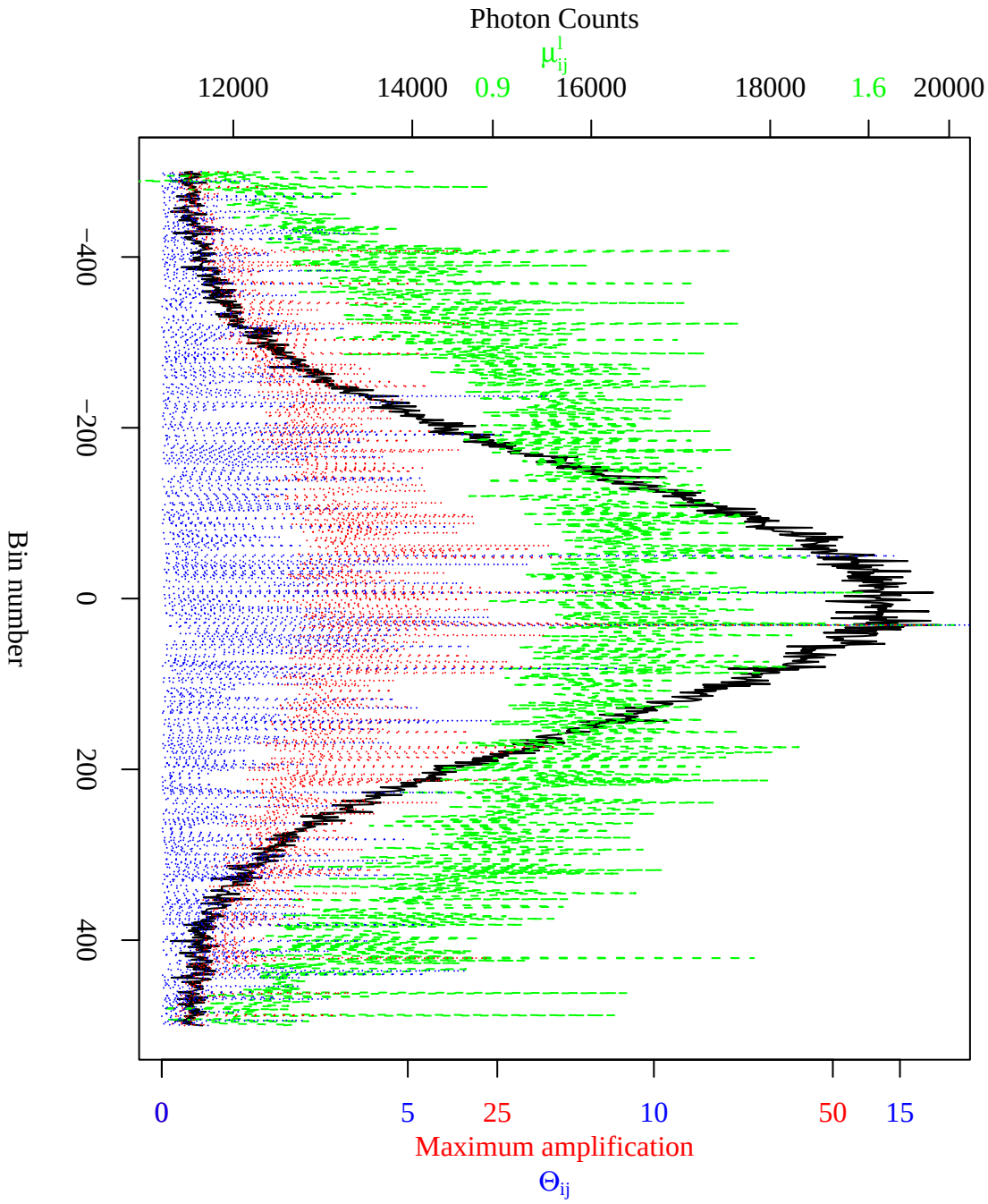

Figure 5.2.: *Photon counts, μ_{ij}^{l}, $\mu_{max,ij}$ and Θ_{ij} for each bin. The line profile in black is produced assuming $N_b = 1$ million clouds, that the line has $EW = 25$ nm, and covers $N_b = 1001$ bins. The $P_d = 0.0025 R_E$ amplification distribution was used. $\mu_{max,ij}$ is in red, μ_{ij}^{l} in green and Θ_{ij} for that bin is in blue.*

While Fig. 5.2 is not as visually appealing as Fig. 5.1, it is possible to discern the important features upon close examination. We see that the blue graph has three high peaks near the line center, and that all these three peaks are associated with fairly high green peaks (though for the leftmost blue peak, there is a higher green peak two bins to the right). In a few cases, we also see that the red maximum amplification peaks correspond to the blue peaks. We see that there are cases were there is agreement between the chosen characteristics and Θ_{ij}.

The correlation coefficients are rather low, 0.27 for $\mu_{max,ij}$ (red) and Θ_{ij} and 0.21 for μ^l_{ij} (green). In Table 5.6, I have again tabulated the correlation coefficients of 10 simulations. As a side note, I saw much higher coefficients in my preliminary analysis, so I may have been particularly "unlucky" with these trials. The graphs of Fig. 5.2 is from trial number 4 in this table.

Table 5.6.: *Correlation between Θ_{ij} and $\mu_{max,ij}$ and μ^l_{ij}. The $P_d = 0.0025R_E$ amplification distribution was used.*

Test number	Θ	Correlation Coefficient	
		$\mu_{max,ij}$	μ^l_{ij}
1	4379	0.15	0.15
2	4117	0.13	0.14
3	4517	0.13	0.14
4	4260	0.17	0.15
5	4344	0.13	0.14
6	4125	0.11	0.12
7	4253	0.09	0.11
8	4469	0.14	0.14
9	4311	0.18	0.18
10	4378	0.14	0.15

Table 5.7.: *Correlation between Θ, the correlation coefficients for $\mu_{max,ij}$ and correlation coefficients for μ^l_{ij}, as tabulated in Table 5.6.*

	$\mu_{max,ij}$	μ^l_{ij}
Statistic	0.21	0.29
$\mu_{max,ij}$	1.00	0.93

We see in Table 5.7 that the correlation between Θ and the μ^l_{ij} is 0.29, which is reasonable for this kind of simulation.

5.3.3. Correlation as a function of maximum amplification

To substantiate the apologetics at the beginning of previous section, I have computed the correlation coefficients for tests made on models with varying maximum amplification. For this, I have used the amplification distribution found in Section 3.2.2, and used the maximum amplification $\mu_{max,an}$ as a free parameter.

Note that in this section, I will speak of "maximum amplification" on two different levels: First of all, the amplification distribution found in Section 3.2.2 sets a fixed limit for the maximum amplification, and this is the free parameter of the distribution, and I use $\mu_{max,an}$ to describe these distributions. Additionally, we are recording the maximum amplification we find in each bin as the simulations proceed, and this is denoted by $\mu_{max,ij}$. This is one of our characteristics, that we correlate with Θ_{ij} for the same bin. The recorded maximum amplification will only in rare cases be as high as the maximum amplification set by the distribution, that is: $\mu_{max,ij} \leq \mu_{max,an}$.

Returning to the details of the procedures of the analysis, I computed 30 tests for each amplification distribution, where $\mu_{max,an}$ was varied from 25 to 200, with increments of 25. The values of the other parameters are as in the previous section.

For each value of maximum amplification $\mu_{max,an}$, I then computed for each test the correlation coefficients for Θ_{ij} and the above described characteristics, in the same way that is displayed in Table 5.4 and Table 5.6. Then, I computed the mean of the correlation coefficients for the tests, and the results are tabulated in Table 5.8.

Table 5.8.: *The mean of 30 correlation coefficients for different $\mu_{max,an}$. The correlation is between Θ_{ij} and $\mu_{max,ij}$, μ^l_{ij} and μ^a_{ij}.*

	$\mu_{max,an}$	Mean Correlation Coefficient		
		$\mu_{max,ij}$	μ^l_{ij}	μ^a_{ij}
1	200	0.363	0.277	0.002
2	175	0.335	0.266	−0.005
3	150	0.312	0.258	−0.010
4	125	0.285	0.246	−0.015
5	100	0.254	0.226	−0.021
6	75	0.224	0.210	−0.029
7	50	0.198	0.190	−0.041
8	25	0.160	0.152	−0.040

We note as expected that the correlation coefficients for $\mu_{max,ij}$ and μ^l_{ij} decline monotonically when $\mu_{max,an}$ becomes smaller. This may justify my contention that it is the characteristics and not necessarily the test that is the reason why the correlation is not very high for extended sources.

Mostly out of old habit, I also computed μ^a_{ij} in these simulations, only to note something surprising: While the correlation between $\mu_{max,an}$ and the other characteristics decline very steadily, μ^a_{ij} *also* decreases very steadily!

At first sight, one might conclude (as I did in Section 5.3.2) that μ^a_{ij} is worthless as a characteristic when the correlation is very close to zero, but the final column of Table 5.8 calls that conclusion into question. It might be that there is some strong relationship between μ^a_{ij} and Θ_{ij}, but that it is of a very non-linear nature. It might mean, in fact, that it is not the test nor the characteristics that should be questioned, but the appropriateness of using the correlation for this particular analysis.

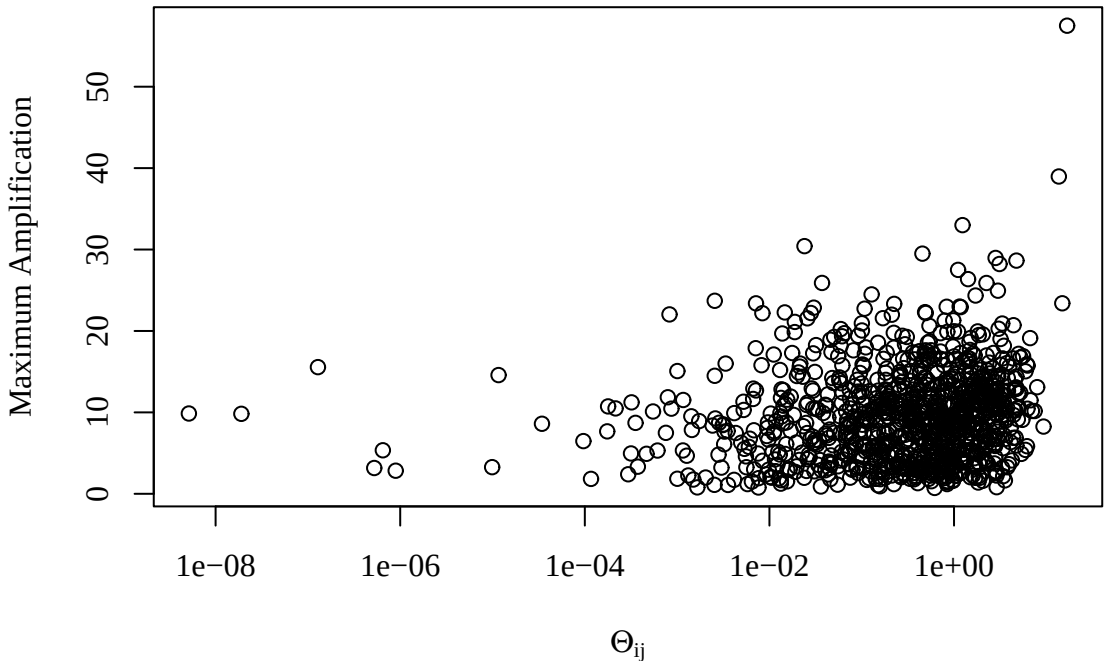

Figure 5.3.: *Pairwise plot of Θ_{ij} and $\mu_{max,ij}$ for $P_d = 0.0025R_E$. The x-axis is logarithmic.*

In Fig. 5.3, in which corresponding pairs of Θ_{ij} and $\mu_{max,ij}$ are plotted, we see that the majority of points are randomly scattered, but there are also a good number of exceptions, which is to be expected from the fact that only a small number of bins are likely to stand out.

However, since I feel I have satisfactorily accomplished what this section set out to do, namely establish that the contingency table based test is capable of identifying the sources of variations, I have not sought to investigate this interesting point further.

5.4. Rejection thresholds

The concept of the *rejection threshold* was introduced in Section 2.4.2 (see page 30). We have noted previously that we are in the fortunate position that both the null hypothesis H_0 and the alternative hypothesis H_1 can be simulated, and therefore, we can compute an optimal rejection threshold as discussed in that section.

I outlined how the quality of the test can be assessed at the end of Section 2.4.2 as well, and stated that assessment is based for a large part on the rejection threshold. Thus, this section will also address that issue.

As in Section 5.3, I have made many simulations, with many tests, but this time with a different purpose: Now, I want to compare tests done on line profiles where lensing is simulated with tests done on line profiles with no lensing.

Specifically, I have used both numerical amplification distributions from Section 3.1, $P_d = 0.01R_E$ and $P_d = 0.0025R_E$, and the analytical amplification distribution from Section 3.2.2 with a maximum amplification of $\mu_{max,an} = 50$, as well as no amplification (i.e. an amplification distribution that has 1 as the only value). Further, I have used parameters of Table 4.1, and made simulations for every combination of the extremes of the free parameters in that table. There are sixteen combinations, and this is done for all four amplification distributions (no lensing included). For every combination, a total of 500 tests are performed, thus I have performed a grand total of 32 000 tests to enable the following analysis.

After computing in R, the results are in the form of huge variable objects that may be many tens of megabytes. Since it is the χ^2-test statistics Θ that is of most interest, I have extracted those into arrays, one array for each parameter set. In this form, the relatively large amount of data is very manageable.

In Fig. 5.4 I have included three histograms of χ^2-test statistics, for three different parameter sets, and with both lensed and unlensed cases plotted in the same histograms. The usefulness of these were pointed out on page 29.

When we examine the histograms of Fig. 5.4, we immediately notice that the test fulfills the most fundamental criterion for a working test: The mean of the test statistics $\bar{\Theta}$ for H_0 (the black histograms) converge to the expected value, namely 3000 for the two upper panels and 300 for the lower. Given the importance of these means in validating the test, I have listed them for all parameter sets in Table 5.9 for reference.

These three examples are chosen for their instructional value: In the upper panel, there is no overlap between the two histograms. In the middle panel, there is some overlap, while the lower histograms overlap substantially. This underlines the points made in Section 2.4.2. Let us first discuss the position of the rejection threshold:

Looking closely at the histograms, one can convince oneself that there is exactly as much of the black histogram on the upper side of the red rejection threshold as of the blue histogram on the lower side. Recall that the probability of concluding “no lensing” falsely is the fraction of the area of the blue histogram on the lower side of the rejection threshold, and probability of concluding “lensing” falsely is the fraction of the area of

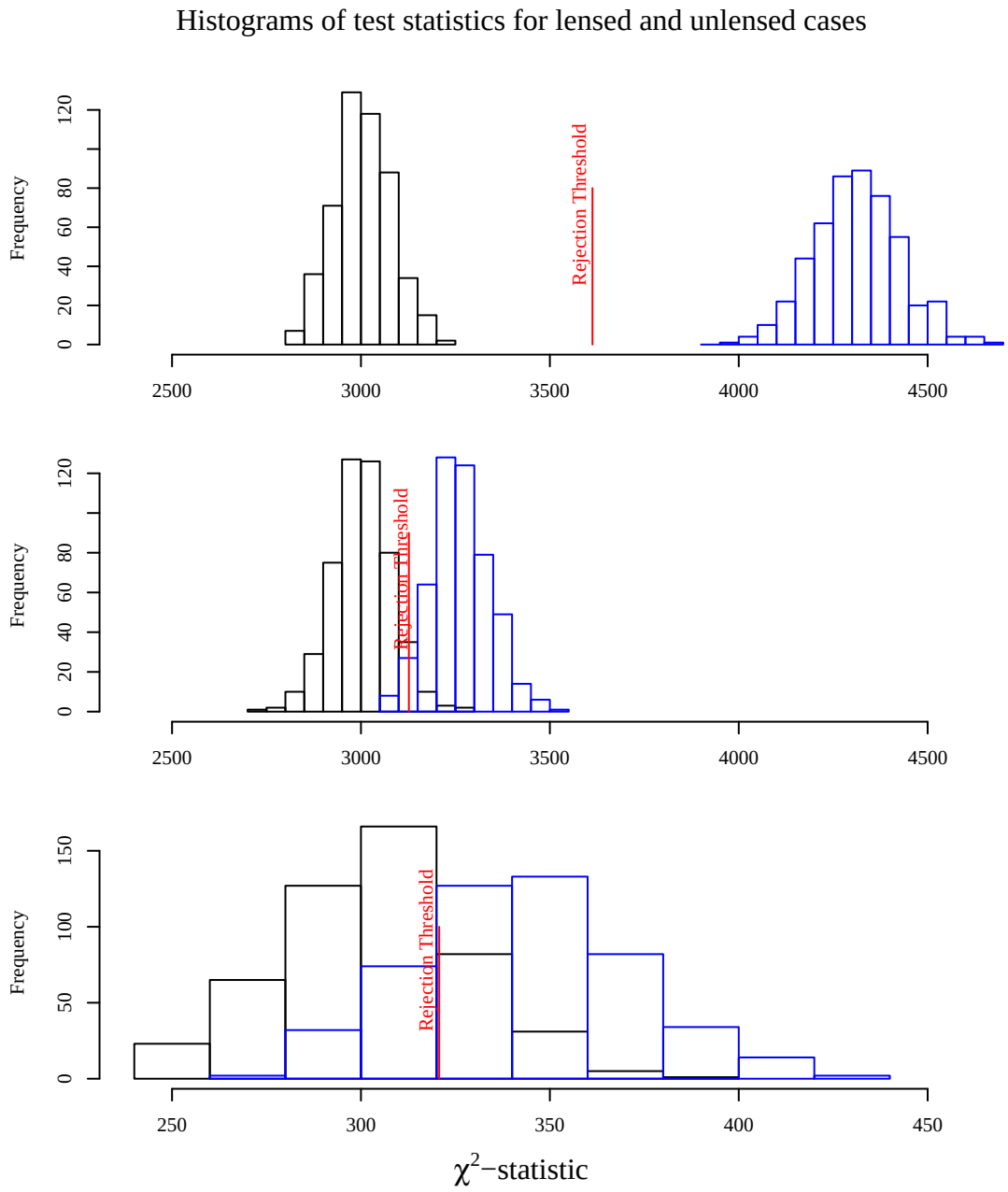

Figure 5.4.: *Histogram of χ^2-test statistics Θ. The black histograms are for unlensed cases, the blue are lensed using $P_d = 0.0025 R_E$. The two upper panels are for $EW = 25$ nm and a thousand bins. The uppermost is for 1 million clouds, the middle is for 5 million. The lower panel is for a case with 1 million clouds, $EW = 8$ nm and 100 bins. Other parameters are as in Table 4.1. The rejection threshold is shown as the red line in each of the panels.*

Table 5.9.: *Mean of test statistics $\bar{\Theta}$ for unlensed clouds.*

	N_c	N_b	EW	$\bar{\Theta}$
1	1000000	8	100	303
2	5000000	8	100	299
3	12000000	8	100	299
4	30000000	8	100	302
5	1000000	25	100	302
6	5000000	25	100	301
7	12000000	25	100	296
8	30000000	25	100	299
9	1000000	8	1000	3003
10	5000000	8	1000	3008
11	12000000	8	1000	2996
12	30000000	8	1000	3002
13	1000000	25	1000	3006
14	5000000	25	1000	3003
15	12000000	25	1000	2998
16	30000000	25	1000	2999

the black histogram on the upper side.

Indeed, we may put the rejection threshold wherever we want. If we push the threshold towards the left, the probability of concluding "no lensing" falsely would decrease, but it would come at the cost of an increase in the probability of concluding "lensing" falsely. Therefore, the rejection threshold now plotted in the histograms can be said to be optimal.

Let us study the panels in more detail: In the first case, there is no overlap, so there will never (for all practical purposes) be any doubt with the optimal rejection threshold, the test will never reject H_0 erroneously. Whenever a test is completed, Eq. (5.2) tells us that if the χ^2-test statistic Θ is higher than the rejection threshold T, in this case 3612, we can safely reject H_0. If it is lower, then H_0 stands firm, and we can be very confident in our conclusion. While this sample does not contain any samples closer to or beyond the rejection threshold, the underlying distributions extends to infinity, and thus the necessity of the "for all practical purposes"-disclaimer.

It is not as well in the second case, as we can only give a probability estimate that we are not mistaken. In this case, $T = 3127$, and therefore the probability of making either error is $\alpha = 0.052$. However, those are good odds (literally!), so, the test is still good.

In the lower panel, the situation has worsened considerably. As noted, there is a large overlap, and much of the black histogram is on the upper side of the red rejection

threshold at $T = 321$ and much of the blue histogram is on the lower side. This means that the probability of making either mistake is $\alpha = 0.2$, and therefore, the test is in this case quite unreliable.

Let us also note what would have happened if we had used a conventional, say 5% level: The rejection threshold for 3000 degrees of freedom is then $T = 3129$, and for 300 degrees of freedom $T = 341$. Comparing with the upper panel shows that it would be a serious mistake to use this level: Our highly reliable test would then have a considerable probability of erroneously report lensing (as an aside, let me mention that this is expected and that it supports the assumption that the test statistic has a χ^2-distribution). For the middle panel, the 5% level is very close to the cited level, so there would have been no difference. In the lower panel, the threshold would have moved right, to the border of the next box. The probability of rejecting H_0 falsely would naturally decrease somewhat, but we see that it comes at a considerable cost, as we would fail to identify lensing in many situations.

5.4.1. If hypotheses should carry equal weight

Note that the choice of optimalization has an additional (and possibly important) benefit: I have taken H_0 to be "no lensing" and H_1 to be "lensing", and stated that it would be a more serious mistake to reject H_0 if it was correct than vice versa. I have done this without much other justification than it being "conventional".

Indeed it is conventional, but we should note the subtleties of the problem at hand: Given a macrolensing galaxy, and given macrolensed images, there would be a fairly high optical depth for lensing, and therefore a conclusion "no lensing" would be no less surprising than the opposite, and therefore, it is not difficult to argue that both hypotheses should carry equal weight, and that in fact, both Type I and Type II errors should be considered equally serious.

I shall not elaborate on this any further, but in fact the present analysis would remain unchanged if one so wished to argue. The reason for this is that the "equal tails" optimalization of the rejection threshold that was chosen in Section 2.4.2 has as one of its features that the probabilities of committing Type I and Type II errors are identical, as we see from the position of the threshold in the histograms.

5.4.2. Type of conclusion

We have now seen three cases where the test goes from highly reliable to unreliable depending on the free parameters of the chosen parameter space and the question is "how do we know we are in either regime?"

The answer is that we cannot tell from this analysis alone, and that was the point in the initial remarks made in this chapter. I have chosen not to answer that *a posteriori* type question: I only state under *what conditions* we can hope to tell that a BLR is micro lensed.

Let us examine the parameters again, they are of a very diverse nature: The number of bins is something we to some degree can control, as the need to obtain better spectra for a particular investigation is something that often prompts technological development. It is also something that is immediately available to us when a spectrum has been obtained.

The Equivalent Width of an emission line is also available to us after observations has been completed, as it can be estimated from the exposures directly. It is however, not something we can control, it is given by the physical processes of the quasar.

Finally, the total number of clouds is neither easily obtainable nor possible to control. It is largely unknown. To be useful, this parameter must be obtained independently.

The type of conclusion that can be drawn from this investigation is primarily: *If* we obtain spectra of an emission line that has an Equivalent Width of 25 nm and covers 1000 bins, and the quasar has 5 million clouds, *then* we have a reliable test and the probability that it will not report lensing falsely when applied to the data is 0.948.

A comprehensive overview of the results will come in the next chapter.

5.5. Simulating more parameter sets

While a huge number of tests has been done in the above, rather few parameter sets have been used. I also need to make simulations for parameters between the chosen extremes. Since for the precision of the statistics, the rule is that "the more, the better", the number of tests done becomes a simple trade-off between the desire for precision and CPU-time available. Clearly, I cannot make as many tests as above for a large number of parameter sets as, it would be prohibitively expensive. Thus, I must lower the standards for how precise I want the χ^2-test statistic to be, and do a smaller number of tests for each parameter set.

To roughly estimate the number of tests that should be done, I made use of the extensive simulations done for the extremes of the parameter space again. For several sets, I merged the same number of statistics from a lensed and unlensed case for the parameter set, and plotted the median for an increasing number of statistics, analogous to what I did with the mean amplification in figures 3.1 and 3.3, to see when they converged so that the fluctuations appeared to be in the second to third digit. While this is very rough, this analysis need not be very rigorous, as, if the number of tests was found to be too small afterwards, new tests could be performed. Based on this, I did a minimum of 150 tests for each parameter set.

6. Results

The core results will be presented in this chapter. I shall give an in-depth look into how the rejection threshold varies with the different parameters, and we shall see under what conditions lensing can be detected reliably, that is, the χ^2-test statistic is higher than the rejection threshold, while at the same time, the rejection threshold itself is higher than a value where we must label the test unreliable.

This chapter is intended to be a container for the results only, the discussion of the results will take place in Chapter 8. Where appropriate, I will point out some important features in the diagrams to be discussed later.

6.1. Extremes of the parameter space

Since I have such a large number of tests for the extremes of the parameter space, I will start with an examination of these tests.

In the tables 6.1, 6.2 and 6.3, I have tabulated the rejection thresholds T, the corresponding significance level α, the means of the test statistic $\bar{\Theta}$ and the p-value for the test[1] for all the combinations of the free parameters. Other parameters are as tabulated in Table 4.1. The different tables are for different choices of amplification distributions. In the first, the amplification distribution characterized by a pixel size of $P_d = 0.01R_E$ is used, in the second, $P_d = 0.0025R_E$, (see Section 3.1) and in the last, $\mu_{max,an} = 50$ is used (see Section 3.2.2).

One conclusion is immediately evident from the tables: It is mostly when there are rather few clouds we can hope to detect lensing under the assumptions we have made.

Let us set our reliability level to $\alpha_r = 0.1$, that is, we will not tolerate a risk of rejecting the null hypothesis of "no lensing" falsely at more than 10%. Using this reliability level, I have extracted the conditions where these simulations substantiate that lensing may be detected from the previous tables, and included them in Table 6.4.

We note as expected that few clouds, strong lines, and many bins are the recipe for detection. Indeed, in the case where this is taken to its extreme, case number 13, the level is nearly immeasurably small. Even in the case for $P_d = 0.01R_E$, where we expect it to be the greatest, it is 2.4×10^{-10}. On the other hand, we note that with a strong line,

[1] Recall that when the rejection threshold is expressed in terms of a probability it is called the "significance level", when the test statistic is expressed in terms of a probability, it is called the p-value.

Table 6.1.: *Rejection thresholds (T), levels (α), mean χ^2-test statistics ($\bar{\Theta}$) and p-values for the $P_d = 0.01R_E$ amplification distribution.*

	N_c	EW	N_b	T	α	$\bar{\Theta}$	p-value
1	1000000	8	100	319	0.217	336	0.157
2	5000000	8	100	302	0.461	306	0.426
3	12000000	8	100	302	0.451	307	0.419
4	30000000	8	100	300	0.491	298	0.524
5	1000000	25	100	331	0.102	374	0.033
6	5000000	25	100	305	0.413	311	0.377
7	12000000	25	100	303	0.433	312	0.366
8	30000000	25	100	300	0.485	305	0.450
9	1000000	8	1000	3095	0.111	3202	0.038
10	5000000	8	1000	3028	0.356	3048	0.328
11	12000000	8	1000	3011	0.443	3025	0.414
12	30000000	8	1000	3005	0.470	3011	0.457
13	1000000	25	1000	3508	0.000	4175	0.000
14	5000000	25	1000	3111	0.077	3234	0.021
15	12000000	25	1000	3048	0.266	3097	0.188
16	30000000	25	1000	3019	0.399	3045	0.340

Table 6.2.: *Rejection thresholds (T), levels (α), mean χ^2-test statistics ($\bar{\Theta}$) and p-values for the $P_d = 0.0025 R_E$ amplification distribution.*

	N_c	EW	N_b	T	α	$\bar{\Theta}$	p-value
1	1000000	8	100	321	0.197	343	0.128
2	5000000	8	100	302	0.463	307	0.425
3	12000000	8	100	301	0.469	303	0.467
4	30000000	8	100	302	0.453	302	0.473
5	1000000	25	100	337	0.068	384	0.017
6	5000000	25	100	306	0.388	314	0.347
7	12000000	25	100	305	0.401	314	0.341
8	30000000	25	100	299	0.499	303	0.465
9	1000000	8	1000	3119	0.063	3234	0.019
10	5000000	8	1000	3026	0.365	3052	0.320
11	12000000	8	1000	3006	0.465	3018	0.434
12	30000000	8	1000	3003	0.482	3006	0.477
13	1000000	25	1000	3613	0.000	4313	0.000
14	5000000	25	1000	3127	0.052	3263	0.008
15	12000000	25	1000	3053	0.244	3112	0.158
16	30000000	25	1000	3019	0.402	3038	0.364

Table 6.3.: *Rejection thresholds (T), levels (α), mean χ^2-test statistics ($\bar{\Theta}$) and p-values for the $\mu_{max,an} = 50$ amplification distribution.*

	N_c	EW	N_b	T	α	$\bar{\Theta}$	p-value
1	1000000	8	100	332	0.098	367	0.040
2	5000000	8	100	304	0.431	312	0.372
3	12000000	8	100	304	0.419	311	0.382
4	30000000	8	100	303	0.440	303	0.468
5	1000000	25	100	352	0.021	429	0.001
6	5000000	25	100	310	0.338	322	0.269
7	12000000	25	100	306	0.390	319	0.305
8	30000000	25	100	301	0.477	306	0.434
9	1000000	8	1000	3181	0.011	3359	0.001
10	5000000	8	1000	3043	0.286	3073	0.246
11	12000000	8	1000	3013	0.430	3028	0.395
12	30000000	8	1000	3011	0.440	3015	0.441
13	1000000	25	1000	3926	0.000	5021	0.000
14	5000000	25	1000	3196	0.006	3410	0.000
15	12000000	25	1000	3080	0.152	3170	0.066
16	30000000	25	1000	3035	0.322	3081	0.236

Table 6.4.: *Conditions where the null hypothesis is rejected, and lensing may be detected for different amplification distributions. Based on the full-size simulations of the extremes of the parameter space. Where H_0 is rejected, the level is quoted.*

	N_c	EW	N_b	$P_d = 0.01R_E$	$P_d = 0.0025R_E$	$\mu_{max,an} = 50$
1	1000000	8	100			0.098
5	1000000	25	100		0.068	0.021
9	1000000	8	1000		0.063	0.011
13	1000000	25	1000	0.000	0.000	0.000
14	5000000	25	1000	0.077	0.052	0.006

and many bins, we may get good results also in the case where the number of clouds is somewhat higher, and that this is in fact the next most favorable condition.

In the case of 12 and 30 million clouds, lensing is not reported with any of the amplification distributions. In row 15 in Table 6.3, the level is nevertheless $\alpha = 0.15$, and that is reason enough to devote some attention to that case in further simulations.

6.2. Denser parameter space

Exactly how the threshold varies with the input parameters is difficult to analyze with so few points, so I have computed a number of observations for a denser parameter space between the limits of the previous. That is to say, for Equivalent Width from 8 to 25 nm, and 100 to 1000 bins. In these simulations, I have, however, not made any simulations for cloud numbers higher than 12 million, since the results of Section 6.1 clearly showed that no lensing can be detected under our stated assumptions for that many clouds. While lensing cannot be said to be detected for 12 million clouds either, it seems still reasonable to simulate for up to that number to see how high it is possible to go.

As stated in Section 5.5 I have performed at least 150 tests for each parameter set. To ensure that I have the highest possible number of tests for each parameter set, the corners of the parameter space were supplemented with the 500 tests done for the extremes of the space. In addition, I had performed a higher number of tests in some cases in my preliminary analysis, and they were included as well. The number of tests for each set does then vary somewhat, and the grand total number of tests for all parameter sets, for the four amplification distributions is 184 352. Yet, fluctuations due to the limited number of tests must be expected.

The current parameter space has three dimensions. It is therefore a challenge to visualize it on two-dimensional paper. The results will come in primarily two forms: Diagrams where the two axes represent two of the three free parameters, and where the level α is shown for all the combinations of the two parameters. The third parameter is the fixed to a given value for a given diagram. The levels themselves are represented by squares with different colors. The tests that reject H_0, that is, where there is a chance of detecting lensing reliably, are mainly marked with a red color, but the color coding starts at yellow for levels which are extremely low and gets a small portion of blue color as the reliability level is approached.

At the reliability level, the color scheme changes abruptly to shades of blue, to clearly distinguish these parameter sets as cases where lensing cannot be detected reliably. For the levels closest to the reliability level, there is a tiny portion of red in the blue color, and on the other end, where the level approaches 0.5, the square will be completely black. Wherever the test is reliable but the test does not reject H_0, the level will be coded by shades of bright magenta, but as we shall see, this situation is very rare. In addition, a contour plot is superimposed on the grid to aid reading. Note,

however, that because of the relatively few values (one for each square), the contours are approximate and do not represent strong boundaries. Finally note that the values on the axes correspond with the center of the squares.

While the color diagrams (hereafter referred to as "images") are designed to give a good view of under exactly what conditions we may detect lensing and how the level varies qualitatively with more than one parameter, I have also made plots that are suitable for studying more quantitatively how the level varies with a single parameter, as well as studying details in the parameter space. These plots have one of the free parameters on the x-axis, and circles marking the levels can be found for each of the sets. The two other parameters remain fixed to a given value. In addition, the p-values are plotted with crosses (though very extreme p-values are in some cases omitted, as their inclusion would have obscured more important features), as well as a dashed horizontal line that signifies the reliability level. Unless stated otherwise, the reliability level used will be $\alpha_r = 0.1$.

Also be aware that the notation (e.g.) 1e-10 that is used on the axes in some of these plots does not refer to the exponential, but means 1×10^{-10}.

Finally, in Appendix B one will find the relevant data for all the sets tabulated where $\alpha \leq 0.125$. The reason why I use this level rather than 0.1 is to enable study of the cases that are just above the reliability level used.

6.2.1. The $P_d = 0.0025R_E$ amplification distribution

First, let us inspect the planes where we expect that the most promising cases will be. In Fig. 6.1, I have included an image that represents the 1 million clouds plane, i.e. the number of clouds has been fixed to $N_c = 1$ million, and the Equivalent Width is shown on the x-axis and the number of bins is on the y-axis. We see that for the combination of a thousand bins and very strong line with $EW = 25$ nm, the level is extremely low, and so coded with an orange, almost yellow color. From that orange corner, the color becomes more red as the number of bins and the Equivalent Width becomes smaller. This is to signify that the level increases for those values. In the lower left end, the color of the squares finally change to blue, since for $N_b = 100$ bins, we find that the level is larger than 0.1 for EW from 8 to 15 nm.

Note that for the three figures 6.1, 6.2 and 6.3, the corners of each image include the large number of simulations described in Section 6.1, and thus the precision is higher for these squares than the other squares.

I have stated an interest in for how many clouds we may still have reliable results. In Fig. 6.4 I have included an image of a plane with $N_c = 5.4$ million clouds as the limit may be in this range, when using $P_d = 0.0025R_E$.

In Fig. 6.5, which is a plane for $N_b = 700$ bins, we note that for a larger number of clouds, the levels have passed the reliability level and that in the image for 100 bins, found in Fig. 6.6, only a few squares are left red. A similar plane for $EW = 18.2$ nm is included in Fig. 6.7.

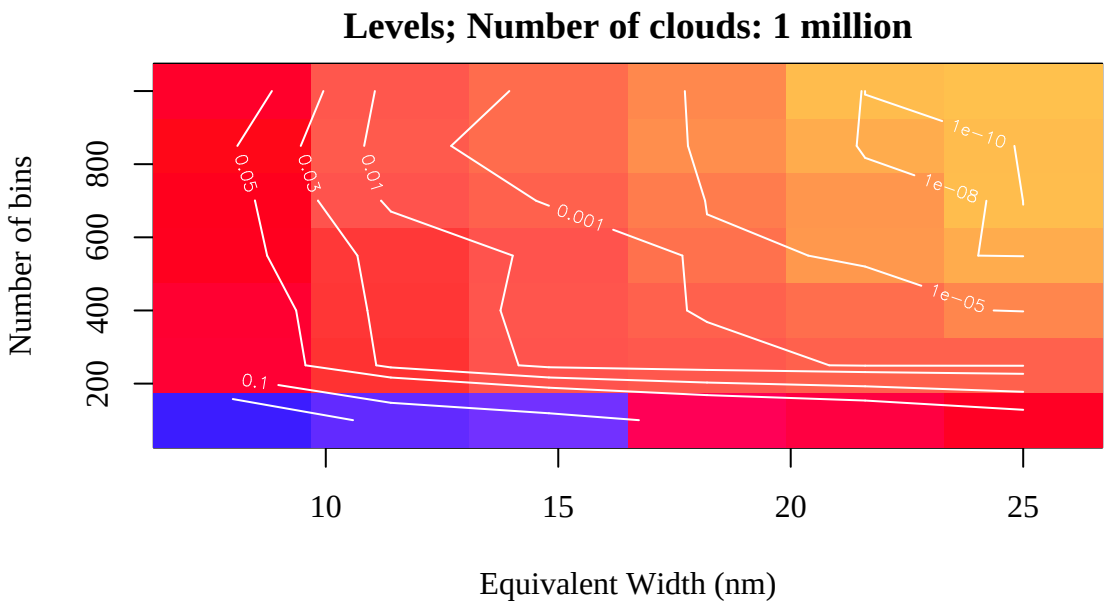

Figure 6.1.: *Significance levels coded in colors and contour lines. Shades of red indicate parameter sets for which the test is reliable, with the most extremely low levels approaching yellow. Tests for parameters sets that are unreliable are indicated with shades of blue. One value per square, and the values on the axes correspond to the center of the square.* $P_d = 0.0025 R_E$ *was used.*

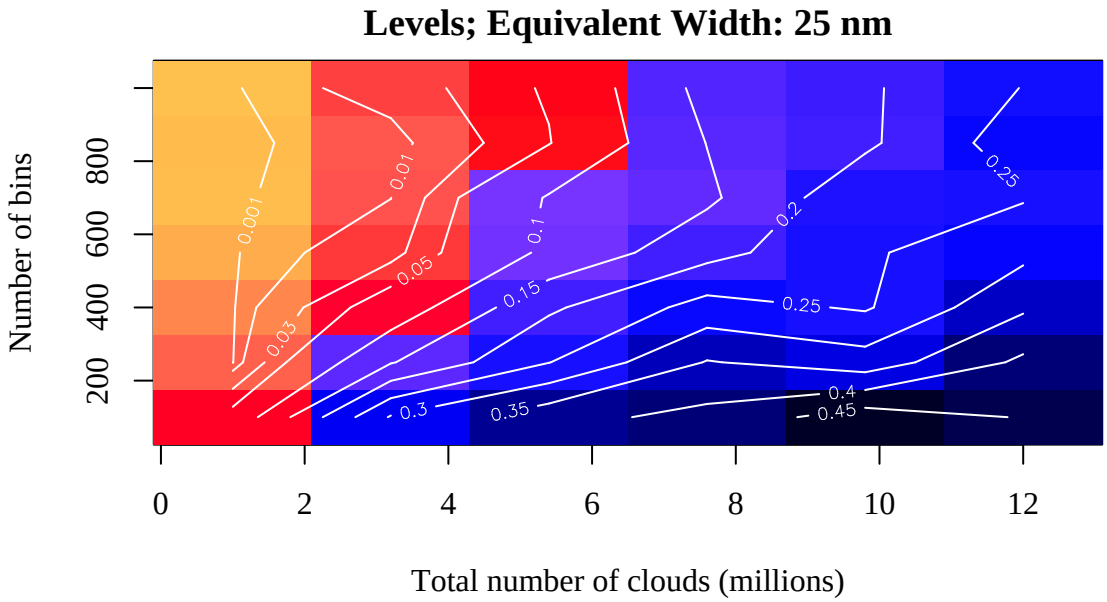

Figure 6.2.: *Significance levels shown similarly to Fig. 6.1. The number of bins and number of clouds are varied.*

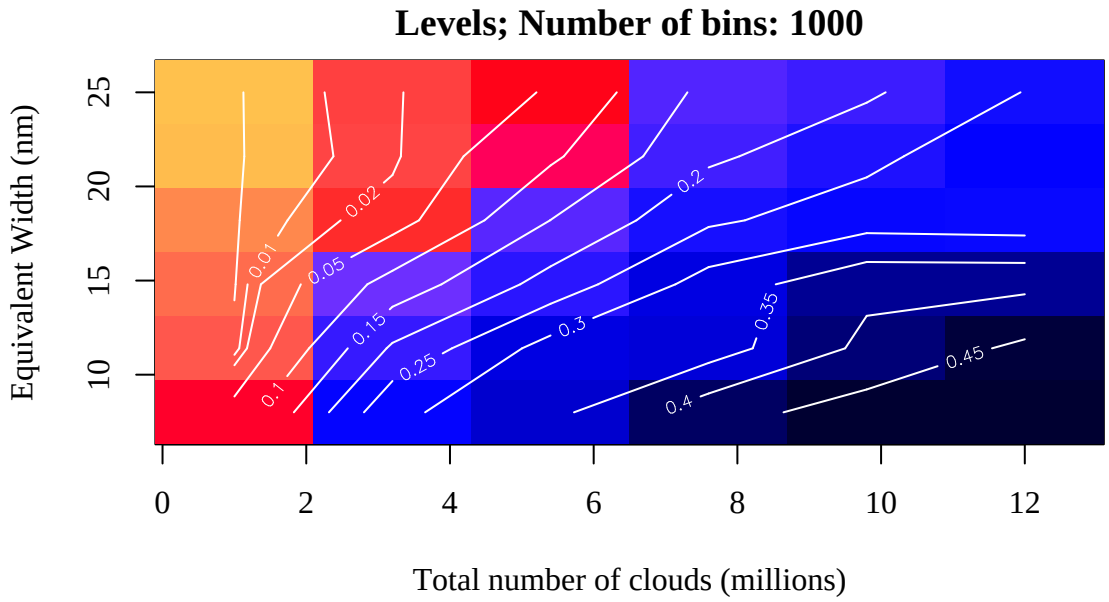

Figure 6.3.: *Significance levels shown similarly to Fig. 6.1. The Equivalent Width and number of clouds are varied.*

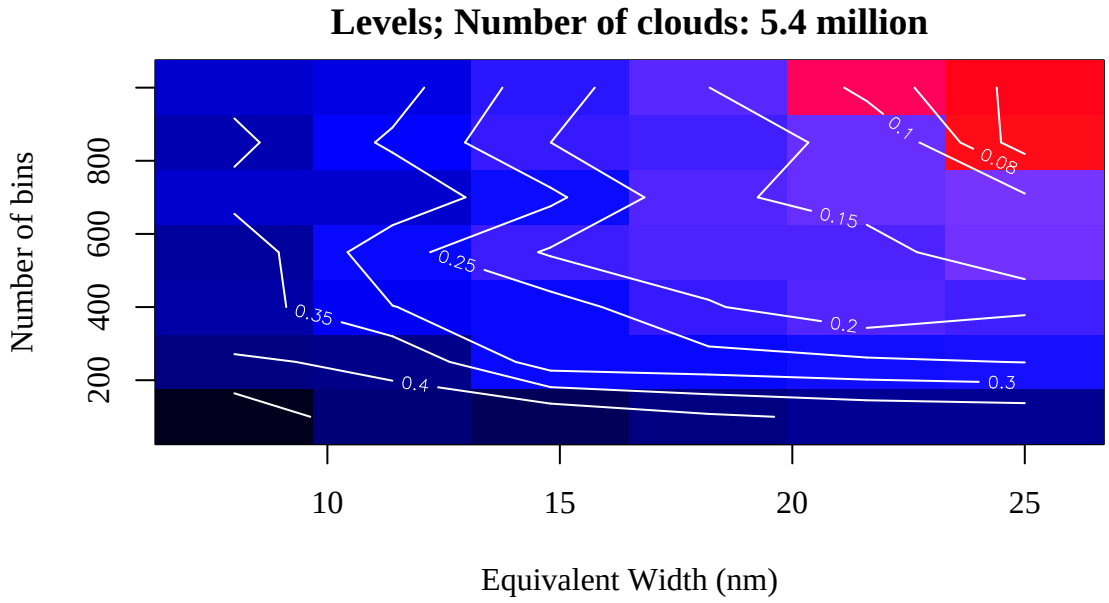

Figure 6.4.: *Significance levels shown similarly to Fig. 6.1. The Equivalent Width and number of bins are varied.*

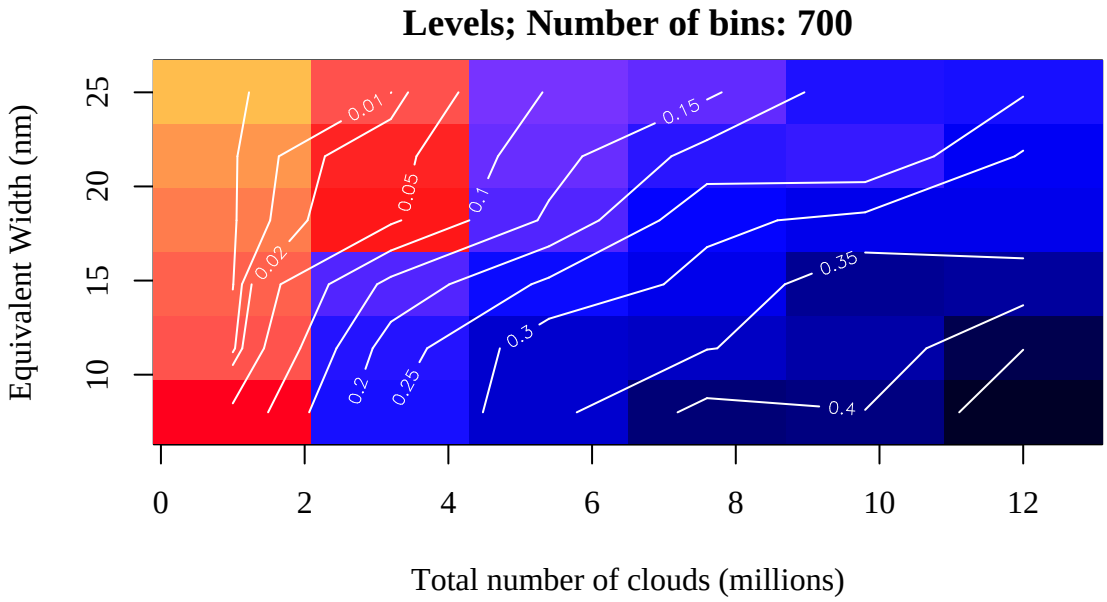

Figure 6.5.: *Significance levels shown similarly to Fig. 6.1. The Equivalent Width and number of clouds are varied.*

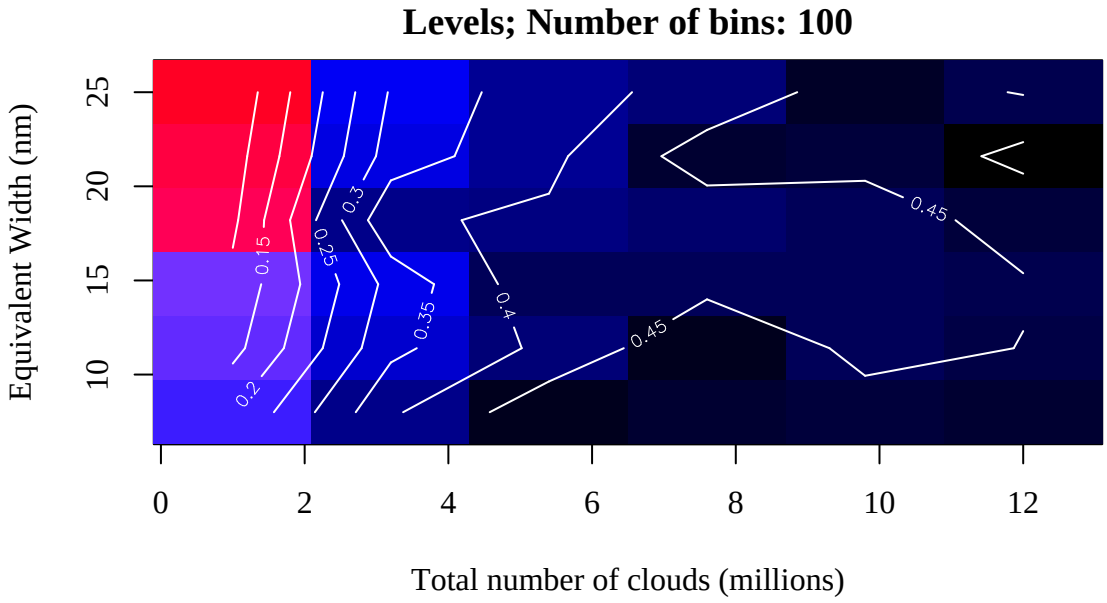

Figure 6.6.: *Significance levels shown similarly to Fig. 6.1. The Equivalent Width and number of clouds are varied.*

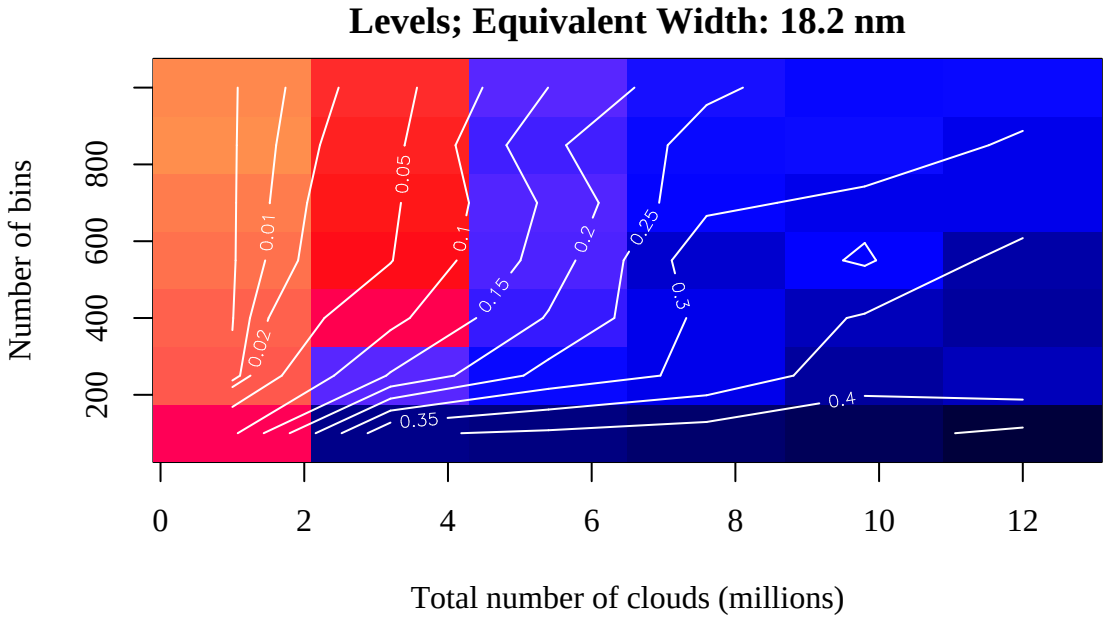

Figure 6.7.: *Significance levels shown similarly to Fig. 6.1. The number of bins and number of clouds are varied.*

In Fig. 6.8, an example of a plot of probabilities is included. The plot shows how the level and the p-value vary with the number of clouds in the case with $EW = 25$ nm and $N_b = 700$, that is, it is a plot along the uppermost squares in Fig. 6.5.

We see that the two first parameter sets come out as reliable, whereas the four last does not, in accordance with the colors in Fig. 6.5. We also see that the p-value approaches the level when the number of clouds increases, but it does not pass the reliability level. It does, however, come very close for N_c =12 million clouds.

In Fig. 6.9, all the points are below the reliability level. Furthermore, it is immediately evident that the logarithm of the levels is roughly linearly dependent on the Equivalent Width, so I made a linear fit and included that as the dot-dashed line.

Under the most favorable conditions, a good fit can also be made for the number of bins, as shown in Fig. 6.10.

The last plot for this amplification distribution shows T and $\bar{\Theta}$. The axes in Fig. 6.11 are *not* logarithmic, but includes a fit where, in fact, the thresholds appears to be linearly dependent on the Equivalent Width.

6.2.2. The $P_d = 0.01R_E$ amplification distribution

We continue the analysis by similarly studying diagrams where $P_d = 0.01R_E$ is used. This means that the clouds are assumed to be larger than in the previous example. In figures 6.12, 6.13 and 6.14, I have included images showing the expected most favorable

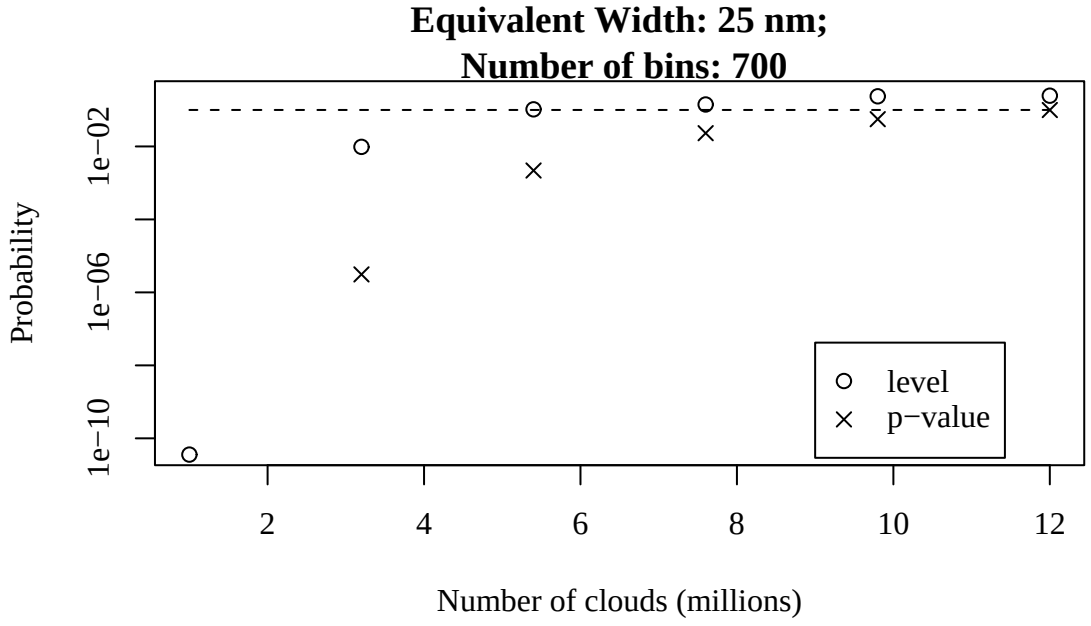

Figure 6.8.: *Significance levels and p-values as a function of N_c for $P_d = 0.0025 R_E$. The dashed line is the reliability level α_r. For instructional purposes, the lowest p-value has not been plotted.*

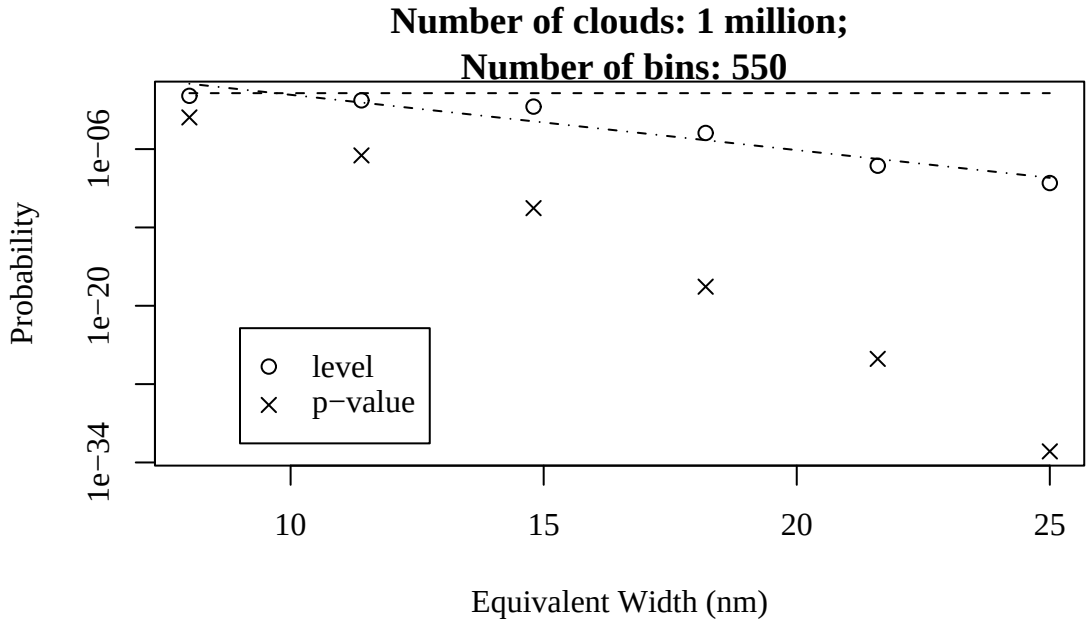

Figure 6.9.: *Significance levels and p-values as a function of Equivalent Width for $P_d = 0.0025 R_E$. The dash-dotted line represents a log-linear fit.*

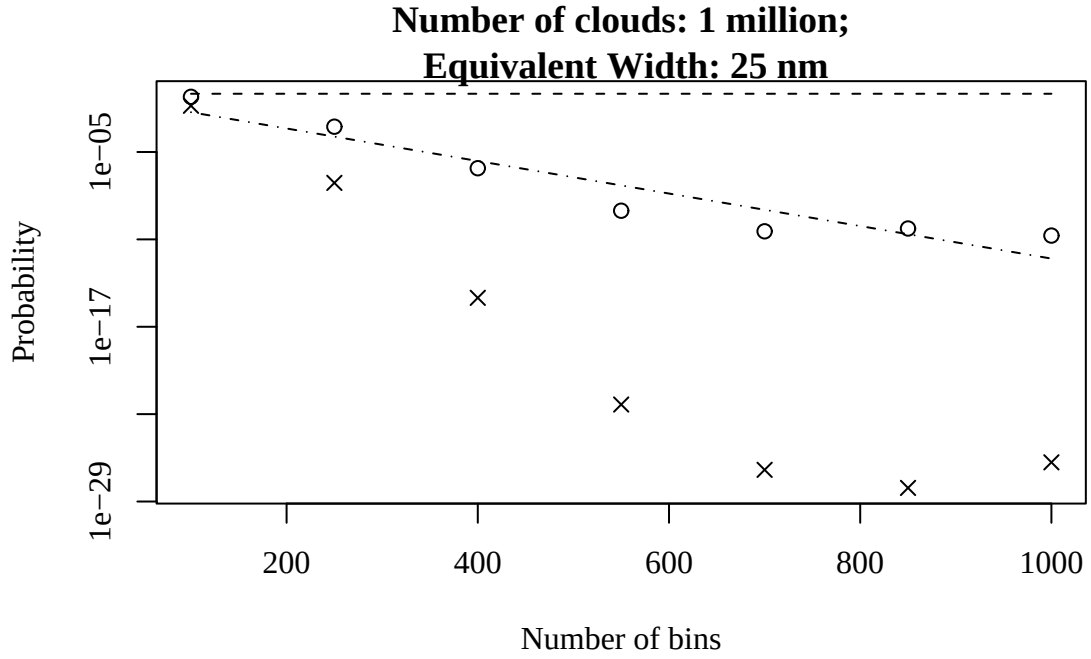

Figure 6.10.: *Significance levels and p-values as a function of N_b for $P_d = 0.0025 R_E$. The dashed line is the reliability level and the dash-dotted line represents a log-linear fit.*

conditions. Again, the value for the corners are the most precise.

We note that for $EW = 8$ nm in Fig. 6.13, the level is rather variable.

Figures 6.15 and 6.16 show where the limit for detection goes, as the former is an image for 5.4 million clouds and the latter is an image for 250 bins.

Take notice of the last point in Fig. 6.17, where the level and the p-value are very close.

In Fig. 6.18, I have again plotted the favorable case where the number of clouds is 1 million and the number of bins is 1000, and again with a log-linear fit. Fig. 6.19 contains a similar plot showing a detail of Fig. 6.15.

Finally, a double-logarithmic plot of T and $\overline{\Theta}$ can be found in Fig. 6.20. A fit has been done where the logarithm of the threshold is modeled as dependent on the logarithm of the number of clouds.

6.2.3. The $\mu_{max,an} = 50$ amplification distribution

Again, the three first figures, 6.21, 6.22 and 6.23, show the most favorable conditions, while Fig. 6.24 shows an image where 9.8 million clouds are used, and therefore represent the least favorable conditions in terms of clouds.

In figures 6.25 and 6.26, two plots where EW and N_c are varied, respectively.

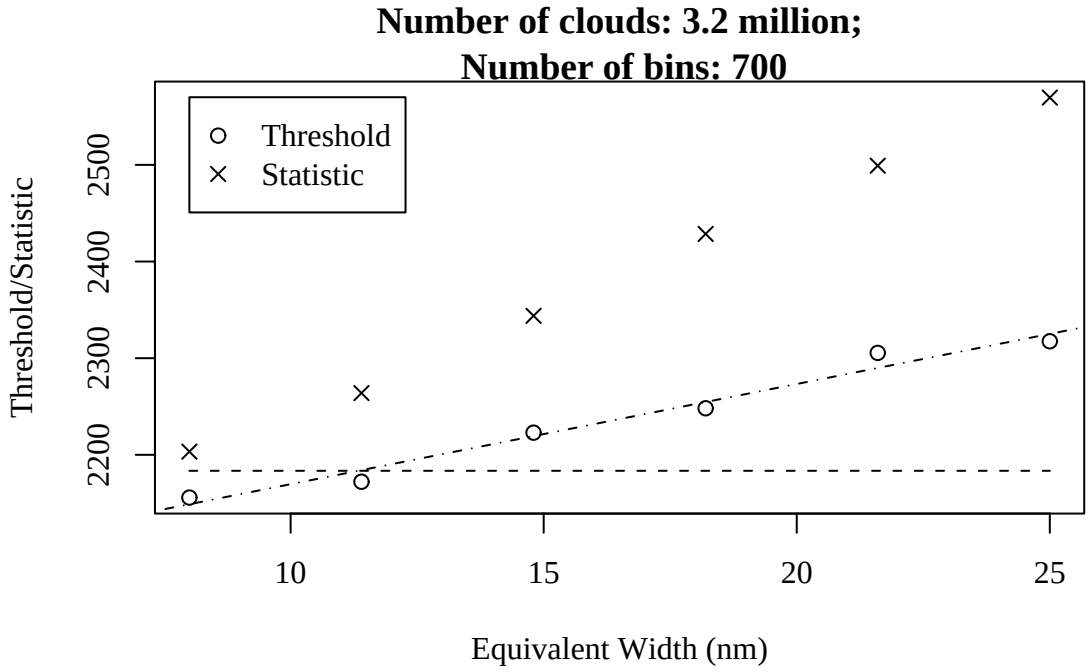

Figure 6.11.: *T and $\bar{\Theta}$ as a function of the Equivalent Width for $P_d = 0.0025 R_E$. The dashed line is the reliability level and the dash-dotted line represents a linear fit.*

The last plot of this section, Fig. 6.27, shows T and $\bar{\Theta}$, but as we can see, the reliability level transformed is not constant. This is because the degrees of freedom in each test is a function of the number of bins, and this makes the use of the test statistics inconvenient for many purposes. Again, a good, linear fit can be made.

6.3. Linear model fitting

In some plots, I have already included fits done by simple least-squares fitting. Let us extend this to making a fit using all three variables at the same time. Since it is reasonable to assume that the points that have been computed using a higher number of tests are more precise, it is reasonable to use a weighted fit. I have done a weighted fit where the number of tests for each parameter set is used as weight. Since we have seen that the threshold itself follows the Equivalent Width (Fig. 6.11) and the number of bins (Fig. 6.27) quite linearly, it is easier to model the threshold than the level. However, we should compensate for the known source of increase in the threshold for increasing number of bins that is due to the increasing degrees of freedom. We can do this by specifying an “offset” term in the model. Finally, while I did a log-log fit in Fig. 6.20, a linear-log fit would also have been good. In the case where several parameters are fitted, it is also much easier to do, and therefore, that is what I have modeled.

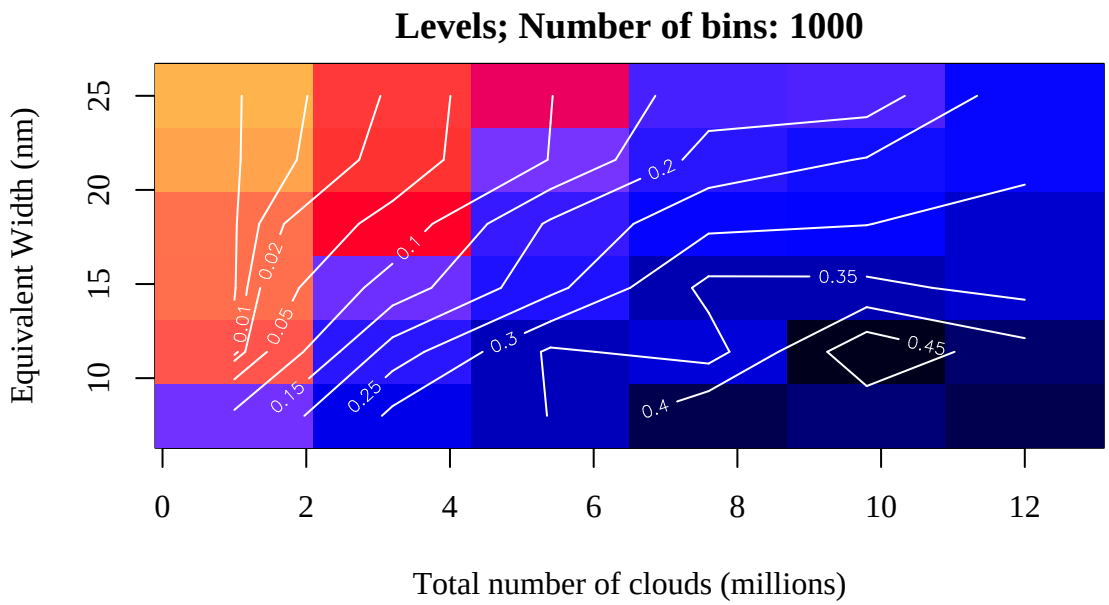

Figure 6.12.: *Significance levels shown similarly to Fig. 6.1, except that* $P_d = 0.01R_E$ *is used. The Equivalent Width and number of clouds are varied.*

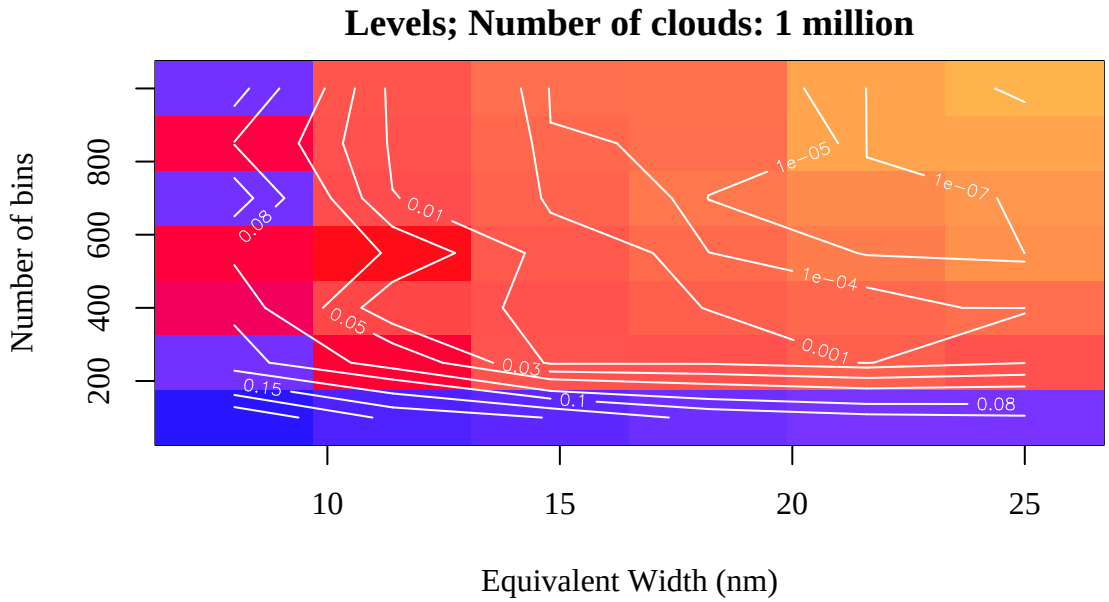

Figure 6.13.: *Significance levels shown similarly to Fig. 6.1, except that* $P_d = 0.01R_E$ *is used. The Equivalent Width and the number of bins are varied.*

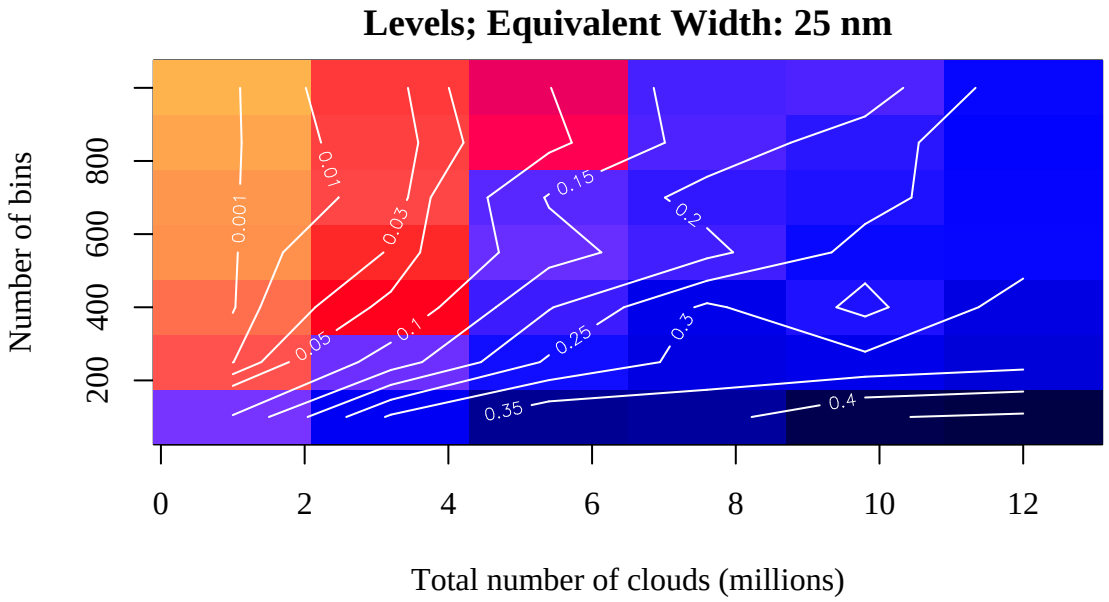

Figure 6.14.: *Significance levels shown similarly to Fig. 6.1, except that $P_d = 0.01R_E$ is used. The number of bins and number of clouds are varied.*

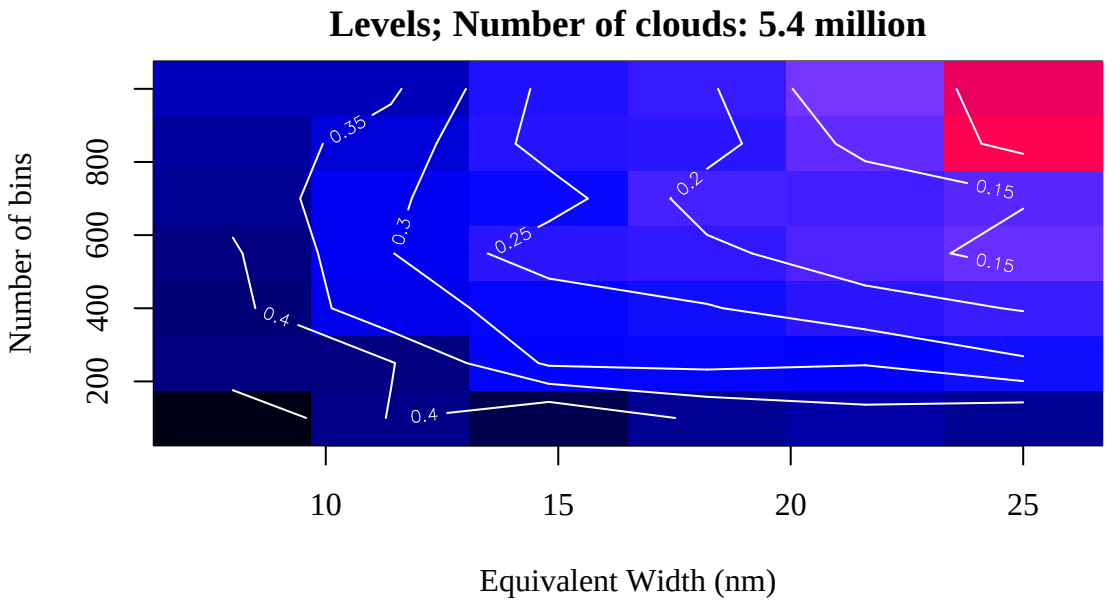

Figure 6.15.: *Significance levels shown similarly to Fig. 6.1, except that $P_d = 0.01R_E$ is used. The Equivalent Width and the number of bins are varied.*

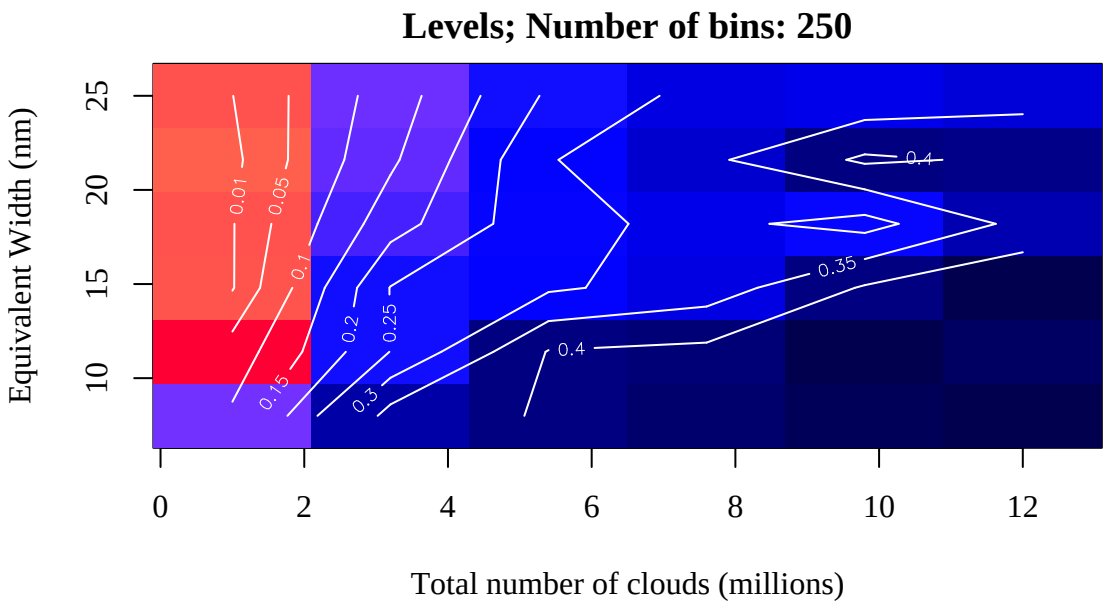

Figure 6.16.: *Significance levels shown similarly to Fig. 6.1, except that* $P_d = 0.01R_E$ *is used. The Equivalent Width and number of clouds are varied.*

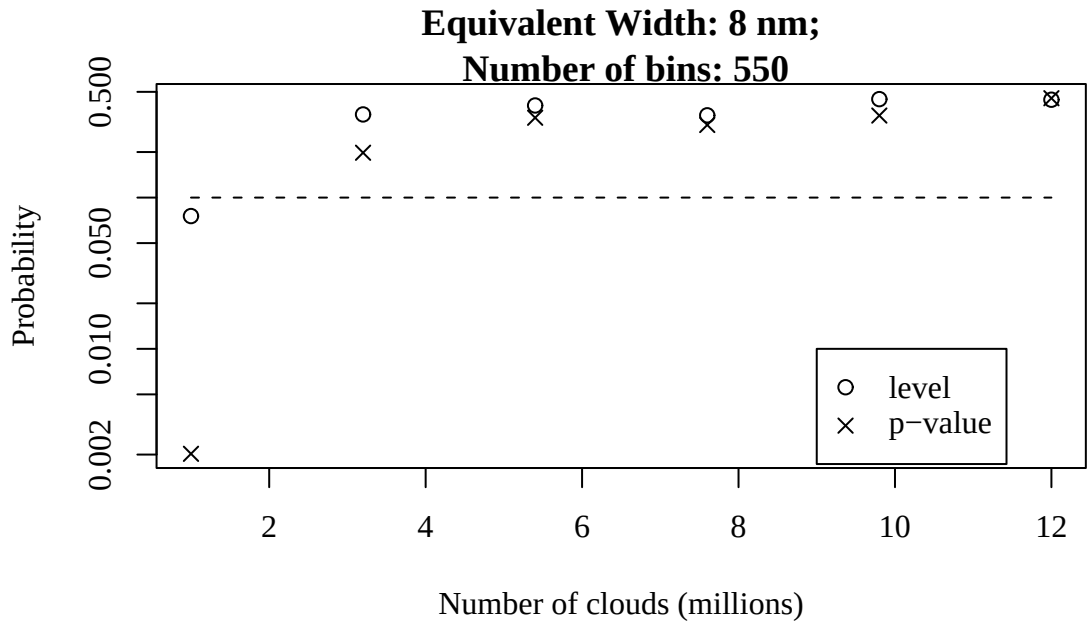

Figure 6.17.: *Significance levels and p-values as a function of* N_c *for* $P_d = 0.01R_E$*. The reliability level is dashed.*

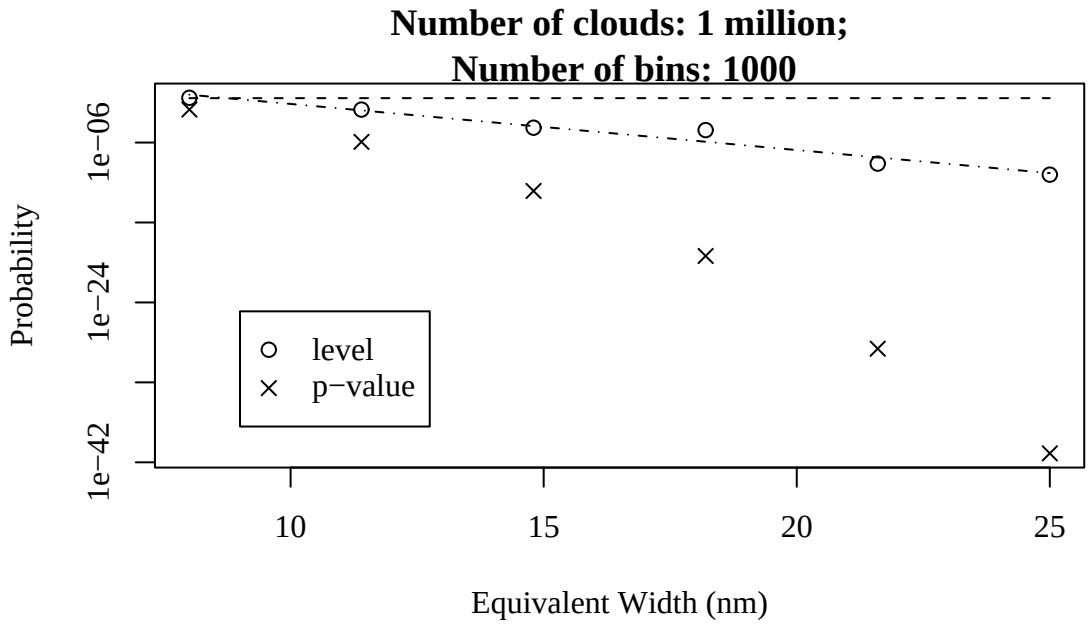

Figure 6.18.: *Significance levels and p-values as a function of EW, similar to Fig. 6.10, but using* $P_d = 0.01R_E$.

When all necessary objects are carefully loaded into R and given appropriate names, the call is straightforward, for example:

```
lx50 <− lm(threshold ~ EW + numberofbins + log(ntotalclouds)
                + offset(3*numberofbins),
                data=lensresx50, weights=numberoftests)
```

This creates an object named lx50 containing details of the fit that can be studied further. In tables 6.5, 6.6 and 6.7, a summary of some characteristics are shown. In addition to estimates of the coefficients of the fit and their respective standard errors, a test has been performed with the null hypothesis that the coefficient is zero. This is a t-test (a t-test was used as an example in Section 2.4.2), and the test statistic and the p-value is quoted. Another characteristic of a fit that is often reported is the so-called "R^2_{adj}". It can be interpreted as the fraction of the variance explained by the model, penalized for the number of parameters. In these cases, R^2_{adj} is 0.9962, 0.9967 and 0.9904 for $P_d = 0.0025R_E$, $P_d = 0.01R_E$ and $\mu_{max,an} = 50$, respectively.

Table 6.5.: *Summary of characteristics of a linear fit for $P_d = 0.0025 R_E$. The coefficients are shown along with their respective standard errors as well characteristics of a test where the null hypothesis is that the coefficient is zero.*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1014.9578	70.0729	14.48	0.0000
EW	4.8423	0.6463	7.49	0.0000
N_b	0.1270	0.0123	10.36	0.0000
$\log N_c$	−71.2872	4.4681	−15.95	0.0000

Table 6.6.: *Summary of characteristics of a linear fit for $P_d = 0.01 R_E$. The coefficients are shown along with their respective standard errors as well characteristics of a test where the null hypothesis is that the coefficient is zero.*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	903.7669	64.4297	14.03	0.0000
EW	4.3225	0.5943	7.27	0.0000
N_b	0.1142	0.0113	10.13	0.0000
$\log N_c$	−63.4699	4.1082	−15.45	0.0000

Table 6.7.: *Summary of characteristics of a linear fit for $\mu_{max,an} = 50$. The coefficients are shown along with their respective standard errors as well characteristics of a test where the null hypothesis is that the coefficient is zero.*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1585.4449	114.6888	13.82	0.0000
EW	7.5698	1.0578	7.16	0.0000
N_b	0.1989	0.0201	9.91	0.0000
$\log N_c$	−111.4363	7.3129	−15.24	0.0000

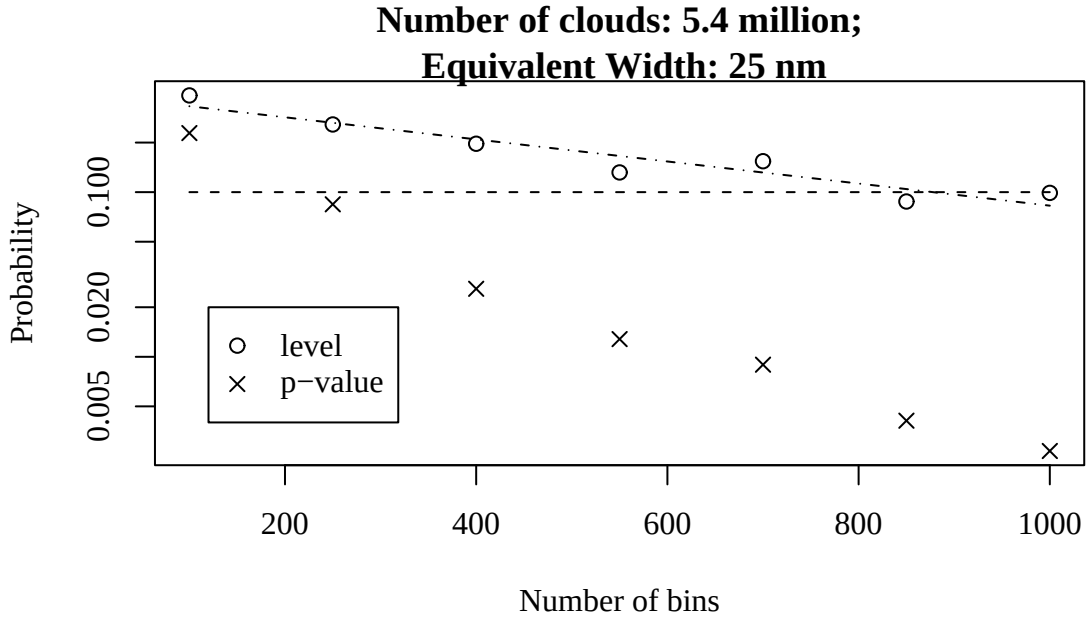

Figure 6.19.: *Significance levels and p-values as a function of N_b, similar to Fig. 6.10, but using $P_d = 0.01R_E$.*

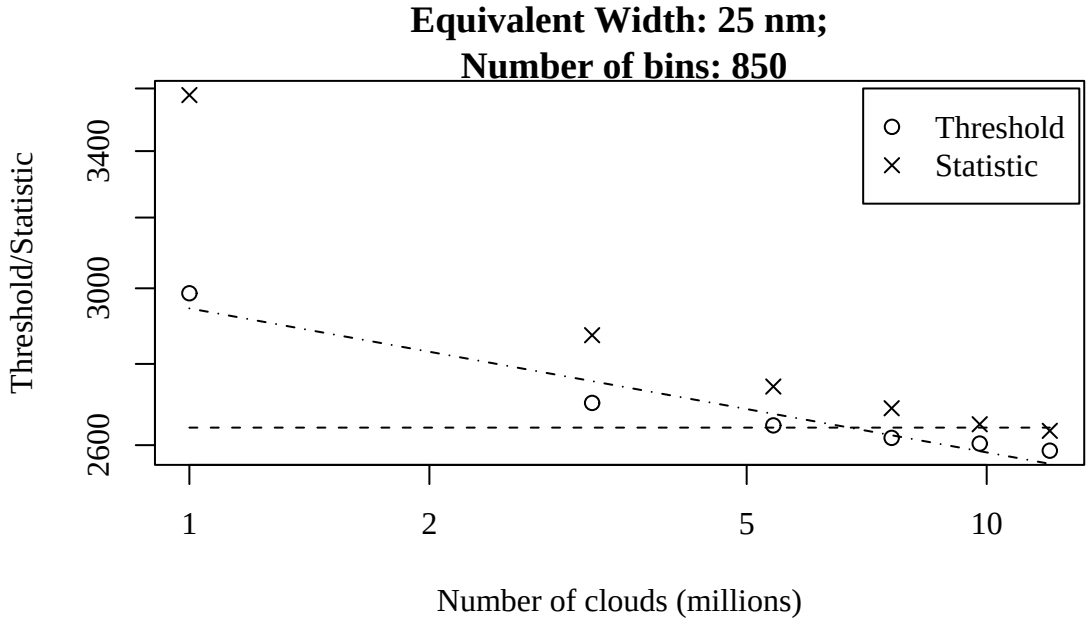

Figure 6.20.: *Threshold and Test Statistics as a function of N_c for $P_d = 0.01R_E$.* Both *axes are logarithmic. The dashed line is the reliability level and the dash-dotted line represents a log-log fit.*

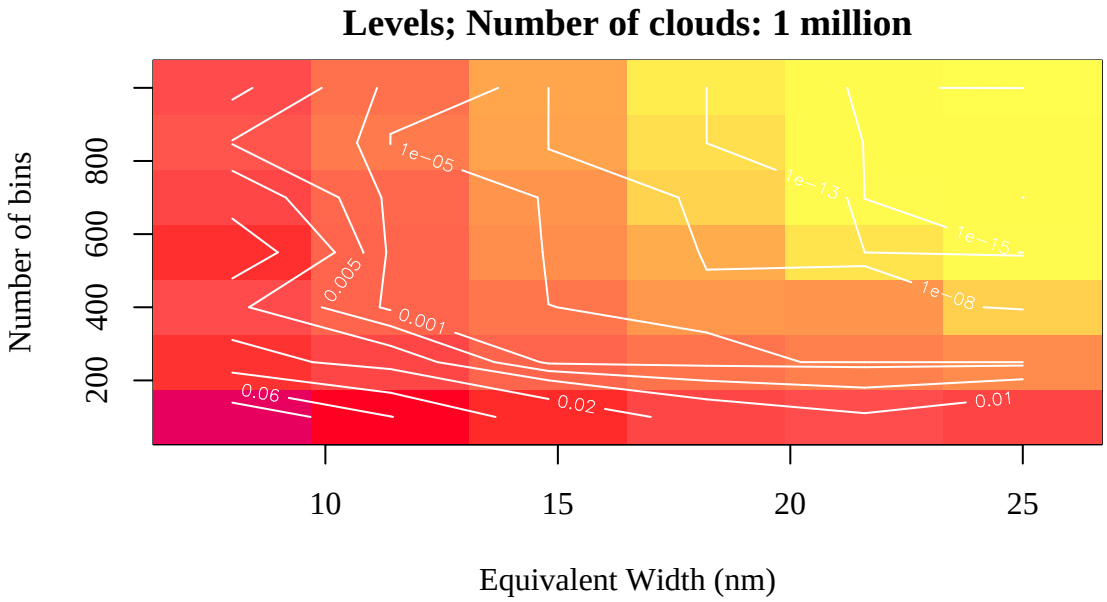

Figure 6.21.: *Significance levels shown similarly to Fig. 6.1, except that* $\mu_{max,an} = 50$ *is used. The Equivalent Width and the number of bins are varied.*

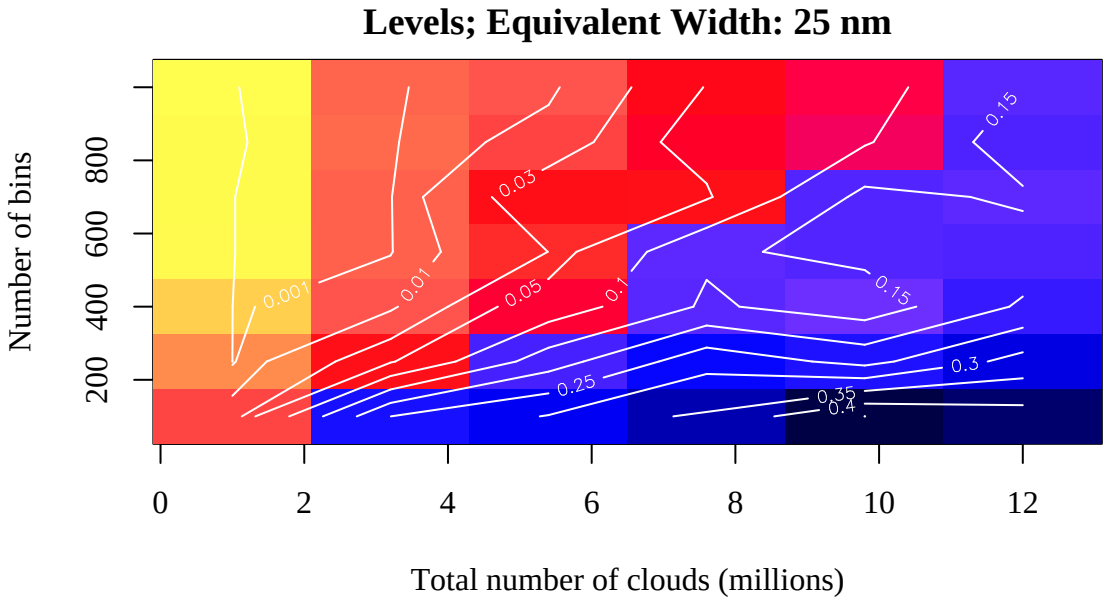

Figure 6.22.: *Significance levels shown similarly to Fig. 6.1, except that* $\mu_{max,an} = 50$ *is used. The number of bins and number of clouds are varied.*

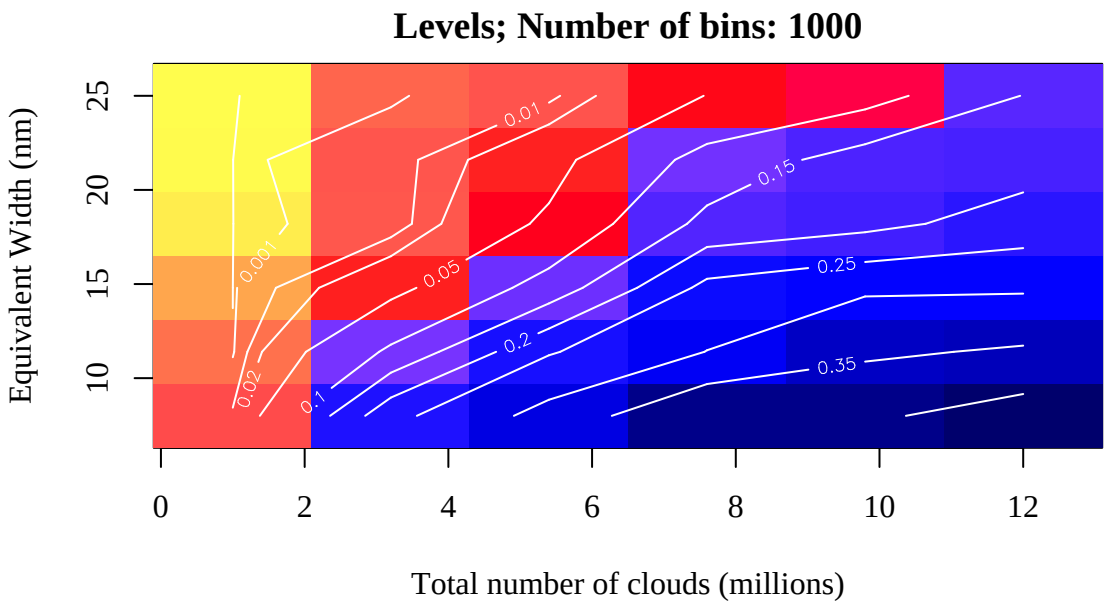

Figure 6.23.: *Significance levels shown similarly to Fig. 6.1, except that* $\mu_{max,an} = 50$ *is used. The Equivalent Width and number of clouds are varied.*

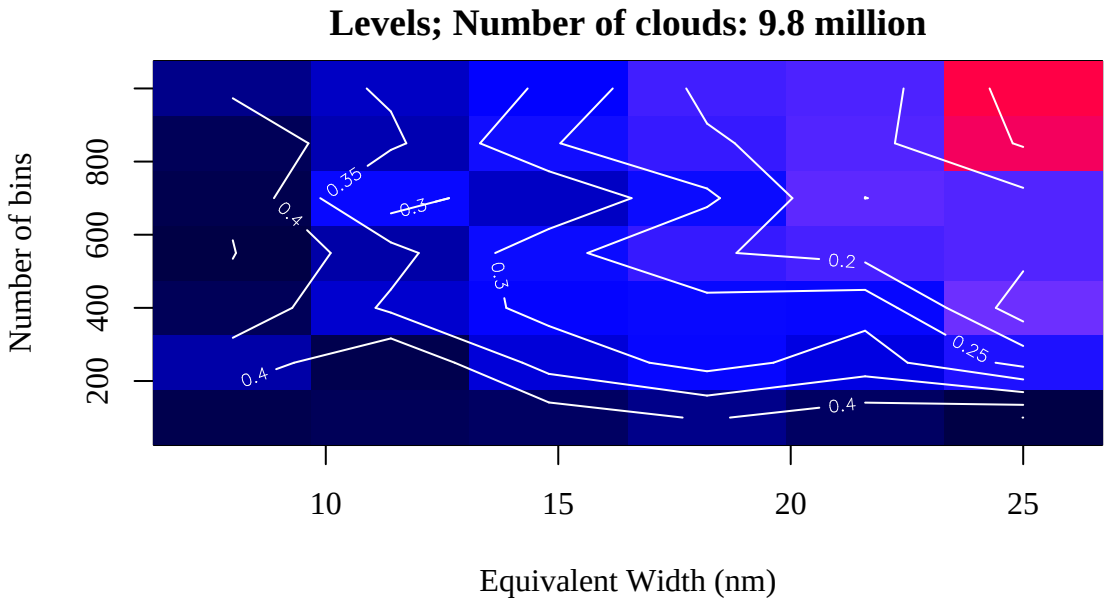

Figure 6.24.: *Significance levels shown similarly to Fig. 6.1, except that* $\mu_{max,an} = 50$ *is used. The Equivalent Width and the number of bins are varied.*

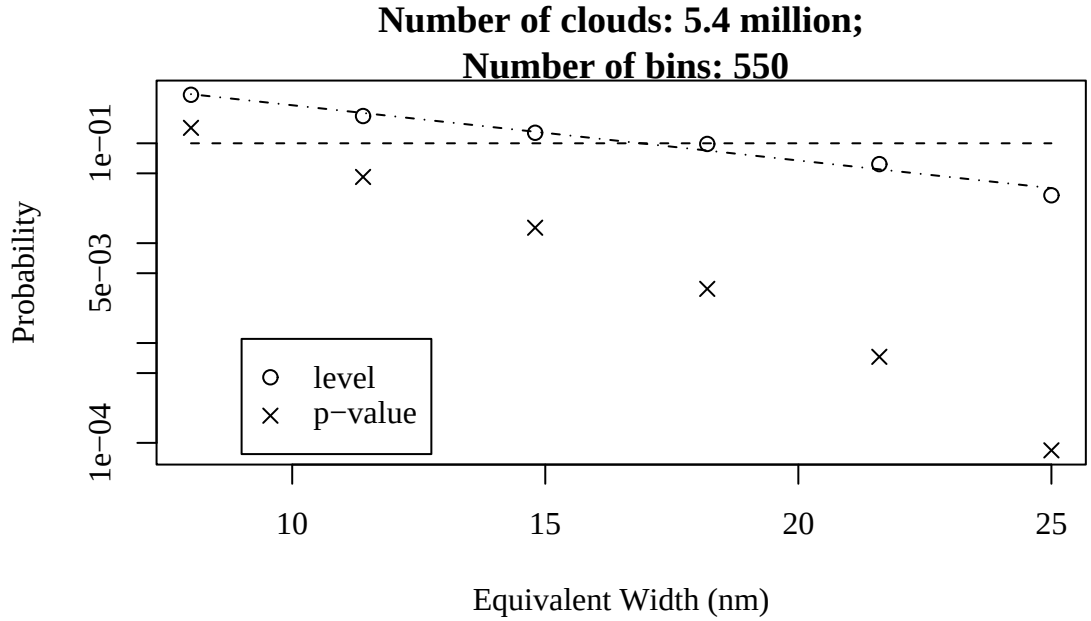

Figure 6.25.: *Significance levels and p-values as a function of Equivalent Width for* $\mu_{max,an} = 50$*. The dash-dotted line represents a log-linear fit.*

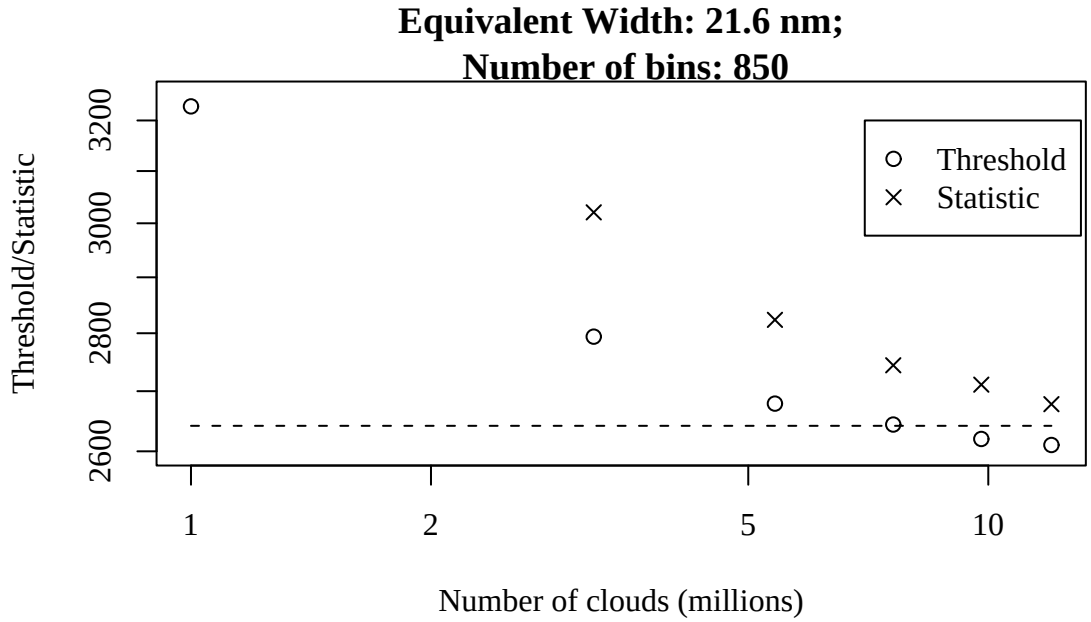

Figure 6.26.: *Threshold and Test Statistics as a function of* N_c *for* $\mu_{max,an} = 50$*.* Both *axes are logarithmic. The dashed line is the reliability level. For instructional purposes, the highest* $\bar{\Theta}$ *has not been plotted.*

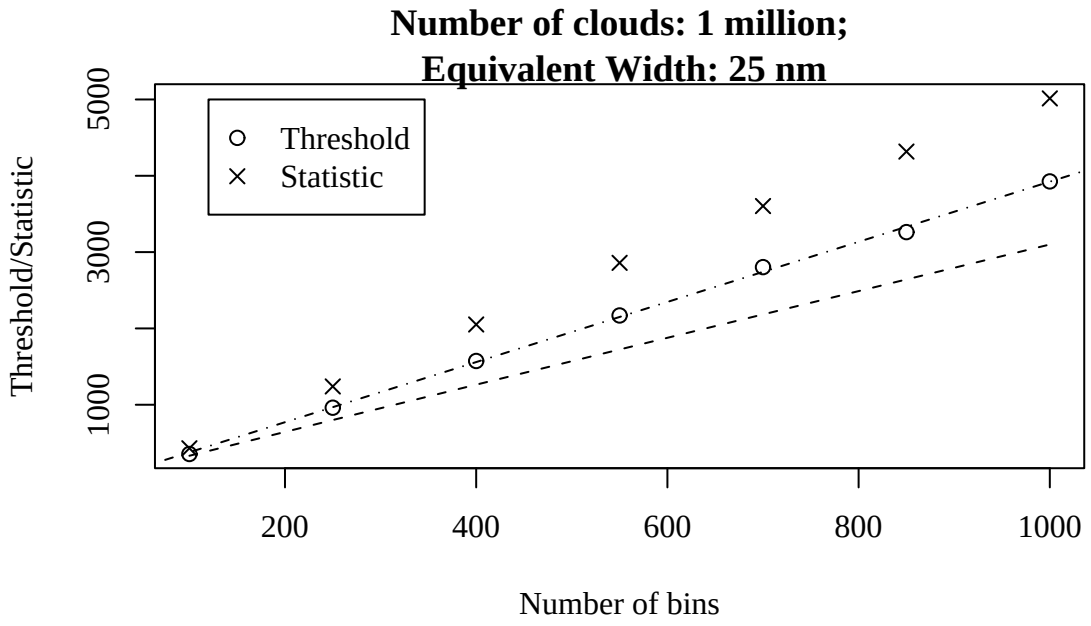

Figure 6.27.: *Threshold and Test Statistics as a function of N_b for a case with $\mu_{max,an} = 50$. The dashed line is the reliability level and the dash-dotted line represents a linear fit. Note that the reliability level is not constant.*

7. Variations

7.1. Level as a function of $\mu_{max,an}$

In the core set, three different amplification distributions were used, two numerical and one analytical where the maximum amplification used were $\mu_{max,an} = 50$. While it is infeasible to compute more numerical amplification distributions, the analytical distribution has the virtue that it is much less expensive in use.

I have varied $\mu_{max,an}$ from 10 to 120 with increments of 10, and computed 580 tests for each. The corresponding unlensed case is the same for all the lensed cases, and therefore, the rejection threshold has been obtained by using the test statistics from the unlensed case as described in the previous chapter.

A plot of T are included in Fig. 7.1. A log-log fit has been done, and illustrated as the dash-dotted line in the plot.

Seemingly, the fluctuations for very high $\mu_{max,an}$ are quite large, so an even larger number of tests would have been advantageous.

I have included a summary of characteristics for the fit in Table 7.1, similarly to Table 6.7 and others.

Table 7.1.: *Summary of characteristics of a linear fit for different $\mu_{max,an}$. The coefficients are shown along with their respective standard errors as well characteristics of a test where the null hypothesis is that the coefficient is zero.*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.9956	0.0042	1902.85	0.0000
$\log \mu_{max,an}$	0.0176	0.0010	16.92	0.0000

7.2. Varying Maximum Number of Counts

In the simulations in the core set, the maximum number of counts were fixed to $N_{max} = 20000$. This is a rather low value, and reflects that we may not be awarded enough observing time to saturate the detector.

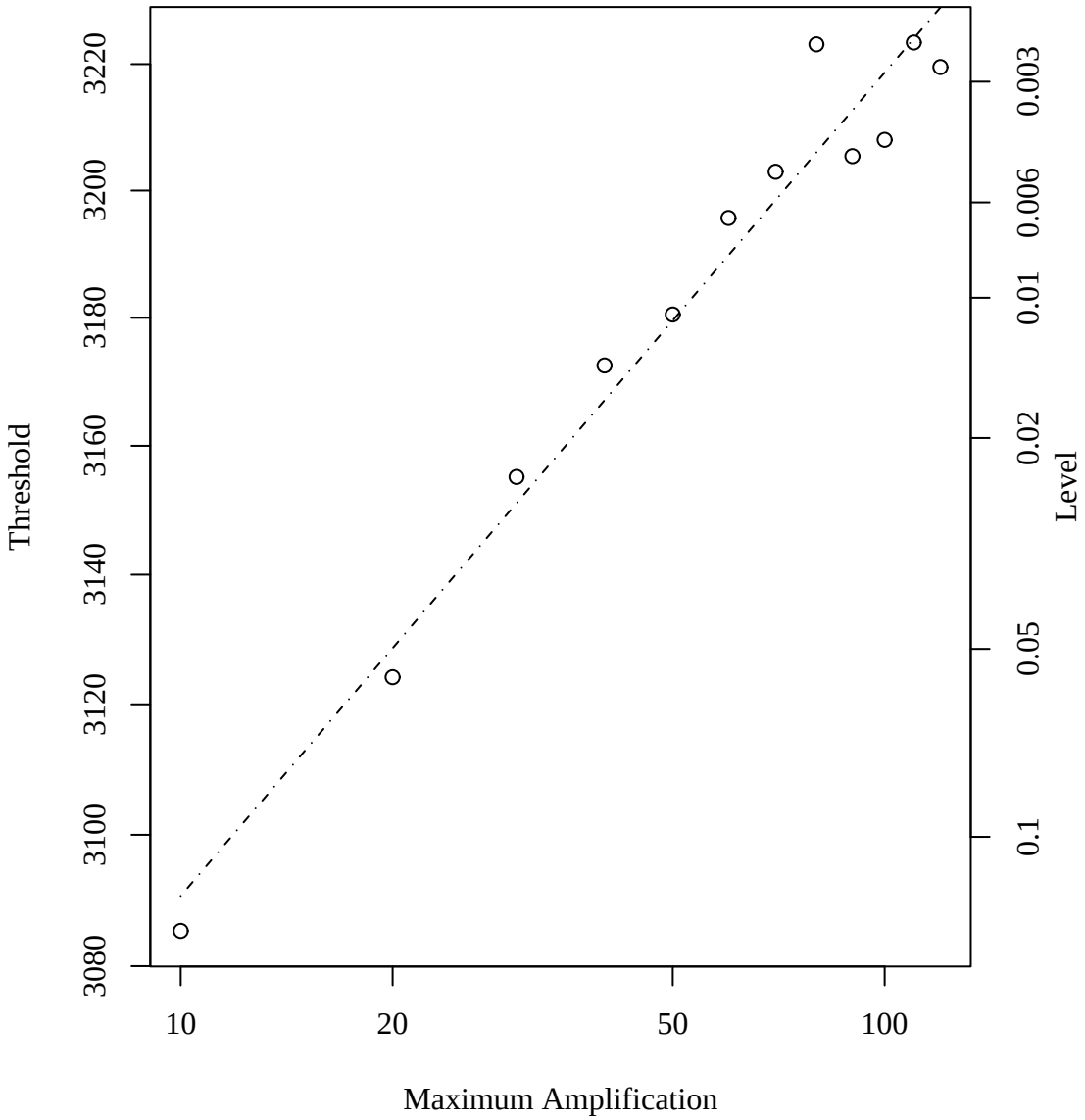

Figure 7.1.: *Rejection threshold T as a function of maximum amplification $\mu_{max,an}$. I have used $N_b = 1000$ bins, $N_c = 1$ million clouds and $EW = 8$ nm. Both axes are logarithmic. The dash-dotted line represents a log-log fit.*

To see if N_{max} influences the results, I have in addition done 500 tests for N_{max} from 40 000 to 100 000 with increments of 20 000 for the unlensed case and for the case with $\mu_{max,an} = 50$. The other parameters were $N_b = 1$ million, $EW = 25$ nm and $N_b = 100$, and further as tabulated in Table 4.1.

First, to be certain that the test doesn't violate the primary criterion for a working test, I checked and found the mean of the test statistics in the unlensed case for $N_{max} =$ 100 000 to be 300.0435, exactly as it should be.

I have plotted the thresholds T in Fig. 7.2, along with a linear fit. Some characteristics of the fit are included in Table 7.2.

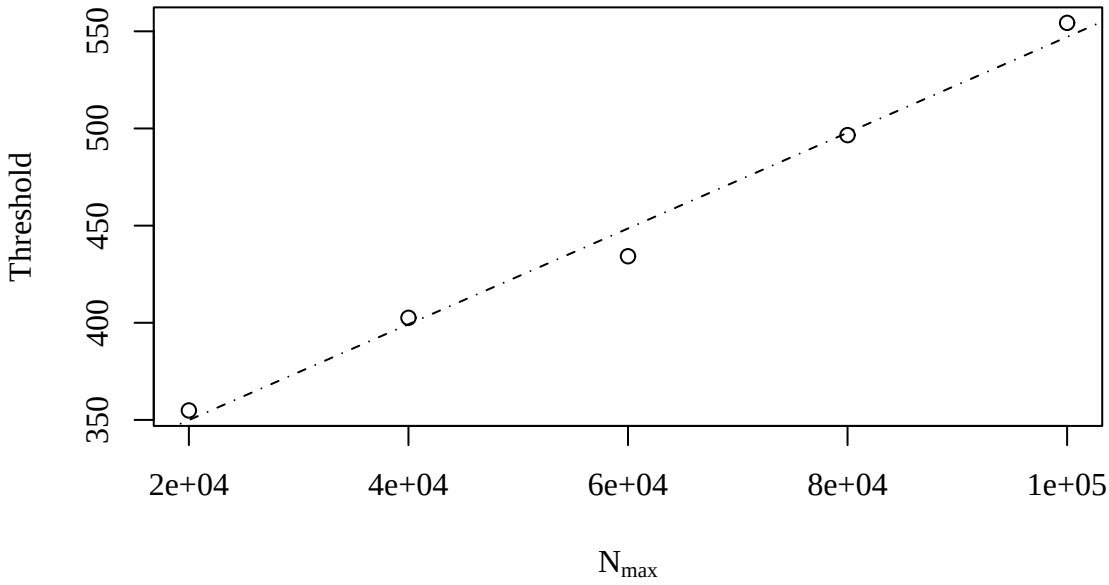

Figure 7.2.: *Rejection threshold T as a function of maximum number of counts N_{max}. I have used $N_b = 100$ bins, $N_c = 1$ million clouds and $EW = 25$ nm. The dash-dotted line represents a linear fit.*

Table 7.2.: *Summary of characteristics of a linear fit for different N_{max}. The coefficients are shown along with their respective standard errors as well characteristics of a test where the null hypothesis is that the coefficient is zero.*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	300.6474	10.3745	28.98	0.0001
N_{max}	0.0025	0.0002	15.76	0.0006

7.3. Two line profiles

While there are many objects such as the Einstein Cross (depicted in Fig. 2.3) that have four images, it is also common that we only see two quasar images.

I have therefore made simulations for two parameter sets with only two line profiles. I have used $N_b = 1000$, $N_c = 1$ million and the $\mu_{max,an} = 50$ amplification distribution. Further, I have used $EW = 8$ and 25 nm, and made 500 tests for both lensed and unlensed cases for these two parameter sets.

First we find that the mean of the test statistics for the unlensed case is 1001.073 and 999.4374 respectively, which is excellent given that the degrees of freedom is 1000.

For $EW = 8$ nm, we find that $T = 1058.8$, corresponding to $\alpha = 0.09597$ and for $EW = 25$ nm, $T = 1289.2$, corresponding to $\alpha = 1.373 \times 10^{-9}$.

7.4. Combining with observations after a delay

The clouds are in motion, and the radial component of the velocity with respect to an observer's line of sight determines which bin the cloud's radiation ends up in. Imaginably, the clouds will sooner or later change their radial velocity so much that their radiation ends up in a different bin than in the current epoch.

Additionally, if the assumption that clouds are lensed independently (Section 4.1.4) is abandoned, the transversal component of the velocity will also be important, because the transversal velocity will then influence at what rate the amplification varies.

After some time, which clouds are in what bin and how much each cloud is amplified is uncorrelated with what is the case in the current epoch.

Exactly how long this time will be, is something that must be computed when a cloud velocity model is assumed, and is thus not a topic for this thesis. I shall only point out that equations (2.36) and (2.37) can be used for this purpose.

The time-delay between the arrival of radiation from the different images (see Section 2.2.1) is also of importance in this context.

Given that the required time has elapsed between epochs, the procedure in Section 4.5 can be run twice to make *two* matrices of observations. Each matrix will then be observations for each epoch.

These can be compared in various ways. As usual in statistics more data means that more reliable conclusions can be obtained.

I have not performed an extensive analysis of this situation, but only checked if the simple contingency table based test could be applied to this situation by simply combining the matrices to one $8 \times N_b$ matrix.

I have computed 500 tests for $N_b = 1000$ bins, $N_c = 1$ million clouds and $EW = 8$ nm. The first criterion for a working test is that the mean of the test statistics in the unlensed case is what is expected under H_0. The degrees of freedom for this test is 7000, and therefore the mean of the test statistics must be close to that value. Unfortunately,

the mean of the test statistics is 7352, which is too high. We must conclude that this simple approach is inadequate.

7.5. Non-monochromatic light

One of the core assumptions of this investigation is that the light from a single cloud is essentially monochromatic, that is, a single cloud contributes all its photons to a single bin (see Section 4.1.2 for details).

I shall outline how this assumption may be abandoned and how an analytical approach may be made to this problem:

Let us for example say that instead of being monochromatic, the photons differ somewhat in their energy, and the number of photons ending up in neighboring bins follows a Gaussian function.

We can then modify the observation matrix O_{ij} to take this into account. Clearly this amounts to convolve the matrix, one-dimensionally along each line profile, with the Gaussian. Then, the expected counts $\hat{E}_{ij}$ according to Eq. (5.1), and then the contributions to the test statistics Θ_{ij} from each bin and image can be computed using Eq. (2.48) and finally Θ is found by summing all elements.

If the Gaussian had a parameter σ that described its width, then Θ will be a function of σ, and we may then compare $\Theta(\sigma)$ with Θ in the case where monochromatic light was assumed, and the impact of varying σ can be studied.

I have not conducted the analysis above, but it is a possible course for further study.

It is obvious, however, that the effect of this is to smooth the line profiles. This, in turn, means that any effects of lensing are weakened, and most likely this will result in that the rejection thresholds will be lowered.

8. Discussion and Conclusions

8.1. Conditions where lensing can be detected

As stated in Section 5.4.2, the main conclusion from the present investigation is to identify the conditions where we can hope to detect lensing.

If we examine the most promising cases that were listed in Section 6.2, mainly the figures 6.1, 6.2, 6.3, 6.12, 6.13, 6.14, 6.21, 6.22 and 6.23, we can extract information that enables us to draw the most important qualitative conclusions.

First, we notice that there is a greater difference between $\mu_{max,an} = 50$ and $P_d = 0.0025R_E$ than between $P_d = 0.0025R_E$ and $P_d = 0.01R_E$. We saw in Section 3.3 that there was good agreement between the numerical and analytical distributions for a range of amplifications. It is however clear that the underlying assumption for the analytical distributions is rather rough, and possibly the numerical amplification distributions are not optimal either.

This urges caution in respect to the precision of the claims made. While the numerical amplification distributions are likely to be a very good representation of the amplification, there is reason to expect that a different choice of amplification distribution will have a significant effect.

It is clear that the size of the clouds is of some importance, as the size of the clouds is an important parameter to any amplification distribution. Indeed, with the assumption that a cloud is equal in size to the pixel size of the raymaps, the cloud size is in practice what distinguishes $P_d = 0.0025R_E$ from $P_d = 0.01R_E$.

For the cases where we have a very large number of simulations (and hence the rejection thresholds have a high precision), it is clear that lensing can be detected under more difficult conditions for $P_d = 0.0025R_E$ than for $P_d = 0.01R_E$. This is summarized in Table 6.4. If we compare for example Fig. 6.3 and Fig. 6.12, we notice that while there are only two squares more that indicate significant results in Fig. 6.3 than in Fig. 6.12 (notably for $N_c = 1$ million clouds and $EW = 8$ nm), the image appears darker. To argue that this is statistically significant, however, would require the use of a statistical test or so many simulations for each set to put it beyond reasonable doubt. Therefore, I rely mainly on the cases where a large number of simulations have been performed when I argue that there is a difference between the $P_d = 0.0025R_E$ and $P_d = 0.01R_E$ based simulations.

Attention is also drawn towards Fig. 6.13 where the level appears to be rather variable for $EW = 8$ nm. While it is not uncommon in the images that the level varies quite a lot for higher values, it is not very often seen around the reliability level. In this case, however, we see that the level changes from being above the reliability level to below and back several times. The change, while not very great, has a visually striking effect since the color scheme changes abruptly around α_r. Most probably, this is *not* significant and due to the limited number of simulations done. As was noted in the text previously, the level in the corners of this figure has been computed by a larger number of simulations, so they are the most precise, and that also supports the contention that these fluctuations are due to the limited number of simulations.

Though the diagrams mostly speak for themselves, the following discussion will lead to the primary conclusions of this investigation:

We have seen that the two amplification distributions $P_d = 0.0025R_E$ and $P_d = 0.01R_E$ can be distinguished from each other. However, we have also noted that the results depend rather strongly on the choice of amplification distribution since these two cases differ even more from the $\mu_{max,an} = 50$-case. This is an important consideration when the primary conclusions are drawn.

My primary conclusions will therefore be of a qualitative nature, and will be based on a conservative evaluation of the main tendencies in the images of Section 6.2.

We see that if there are 1 million clouds, we have usually a very good chance of detecting lensing, as seen from figures 6.1, 6.13 and 6.21. While one should seek to find strong lines and an instrument with good resolving power, unavailability of this is insufficient reason to not make observations if independent studies have argued that quasars have few clouds.

If there is little or no knowledge as to the number of clouds, having a very strong line and a detector that can resolve the line into as many as $N_b = 1000$ bins become desirable. We see from e.g. Fig. 6.5, Fig. 6.7, Fig. 6.12 and Fig. 6.16 that detection is very difficult for small values of N_b and EW, and therefore, one should choose instruments and objects that can provide high values for these parameters in the preparations for an empirical investigation.

8.2. Full linear model fitting

In Section 6.3, I made a linear regression where T was the response variable and N_b, N_c and EW were covariates. The characteristics reported were truly excellent. Visually, the simple fits included in the figures 6.11, 6.20 and 6.27 are undoubtedly appealing. Nevertheless, we must ask if this is all too good to be true.

The simple linear modeling done depends heavily on the assumption that the errors are normally distributed. If this is not the case, the conclusion drawn may have a weak foundation. A common approach to make a first assessment of the validity of this assumption is to make a so-called “Q-Q plot”, that is, we plot the standardized residuals

as a function of the theoretical quantiles. The important point is that if the residuals are indeed normally distributed, the points will be on a straight line. A plotting routine for this analysis is a part of R's standard tools for the study of linear regression objects. One such plot is included in Fig. 8.1.

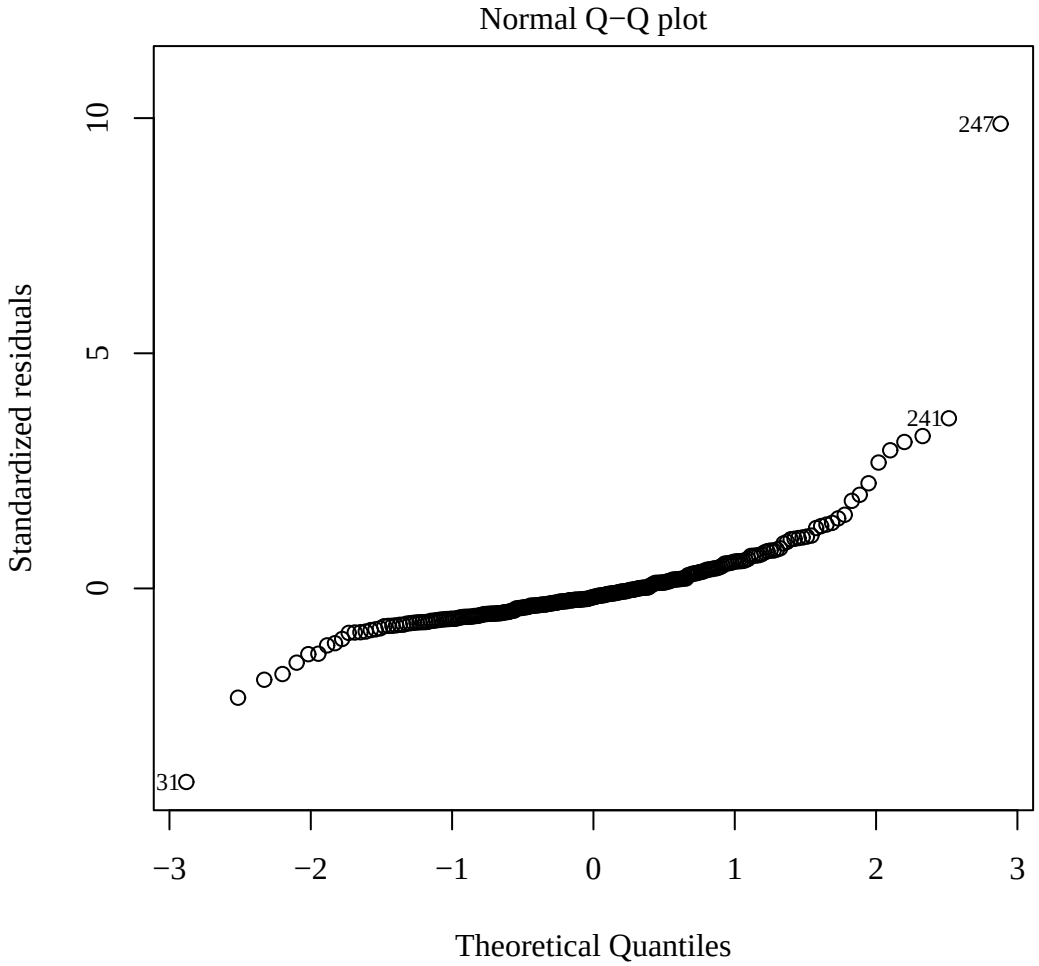

Figure 8.1.: *Q-Q plot for the residuals of the fit described in Section 6.3 for* $\mu_{max,an} = 50$*. If the errors are normal as assumed, the points would lie on a straight line. The numbers in the plot is the number of the parameter set that the "outlier" stems from.*

We see that the residuals are quite far from being normal. Since the variable under consideration is the median of a (approximate) χ^2-distribution, this is not surprising.

R suggests another important plot to use for the first assessment of the adequacy of a linear model: Plotting the residuals as a function of the fitted values. Such a plot is found in Fig. 8.2.

If the real relationships had been found, we would expect to see no structure in this plot, only random variations around 0. However, we clearly see structure. While

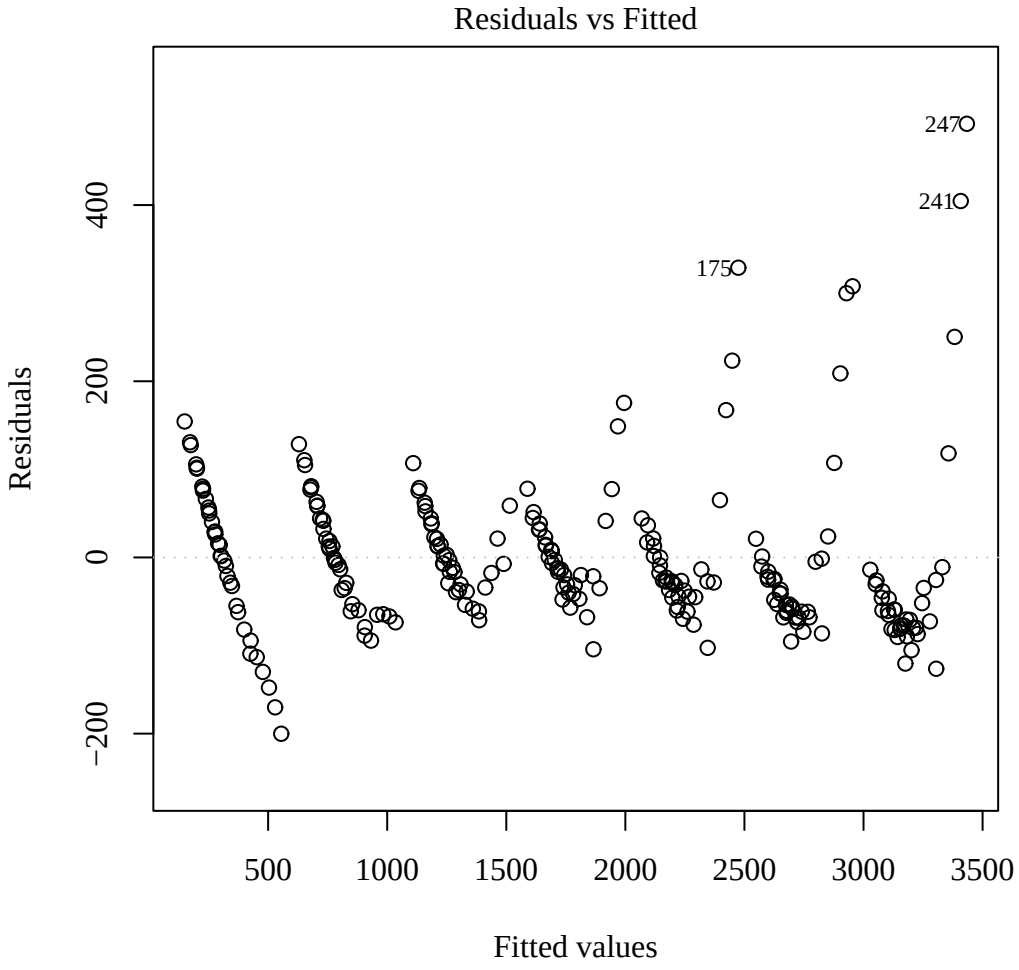

Figure 8.2.: *Residuals as a function of fitted values for* $\mu_{max,an} = 50$*. The numbers in the plot is the number of the parameter set that the "outlier" stems from.*

the above plots are only for the $\mu_{max,an} = 50$ amplification distribution, plots for other amplification distributions are similar.

Clearly, there are problems with the linear regression shown in Section 6.3. The first problem is that the residuals are not normally distributed. This means that, among other things, the standard errors quoted in tables 6.5, 6.6 and 6.7 do not accurately describe the uncertainty in the coefficients. This is a problem that may be addressed by using "Generalized Linear Modeling" (GLM), that allows a wider range of response distributions. However, to find the real relationship between the parameters and the rejection threshold, it is likely that non-linear modeling must be employed.

This does not mean that the linear fit is without value, it only means that caution must be exercised when using the results. For prediction, that is, finding values other than those simulated in or close to the range simulated, the analysis is probably adequate.

For conclusions about the exact nature of the causal relationship between the number of clouds, the Equivalent Width, the number of bins and the threshold (and thus the level), one must be more modest. We may only say that the result is indicative of the relationship.

Thus, with the reservations mentioned above, we may draw the following secondary conclusions:

- For the case of the $P_d = 0.0025R_E$ amplification distribution, the rejection threshold T for a test can according to Table 6.5 be represented by the following equation:

$$T = 4.84EW + 0.127N_b - 71.3\log(N_c) + \mathrm{df} + 1015, \tag{8.1}$$

  where df is the degrees of freedom for the test in question.

- For the case of the $P_d = 0.01R_E$ amplification distribution, the rejection threshold T for a test can according to Table 6.6 be represented by the following equation:

$$T = 4.32EW + 0.114N_b - 63.5\log(N_c) + \mathrm{df} + 904. \tag{8.2}$$

- For the case of the $\mu_{max,an} = 50$ amplification distribution, the rejection threshold T for a test can according to Table 6.7 be represented by the following equation:

$$T = 7.57EW + 0.199N_b - 111.4\log(N_c) + \mathrm{df} + 1585. \tag{8.3}$$

8.3. P-value higher than level

I mentioned in the description of the images that cases where the p-value was higher than the level but lower than the reliability level would be marked with shades of bright magenta, but none of the images contained any such squares. Indeed, none of the

simulations done contained any such cases, but there is one single case where the p-value was higher than the level, and this case is shown in Fig. 6.17.

It is not surprising that this situation is very rare, and certainly not that it has not happened when the level is below the reliability level. If we return to Fig. 5.4, we see that for this to happen, the rejection threshold, which is a function of the test statistics for both the lensed and unlensed cases, must be pushed to the left of the mean of the test statistics for the unlensed case. That is something that is likely to happen only if the test statistics for the lensed case are overall no greater than for the unlensed case. Thus, it can only happen when the threshold is close to 0.5 (provided that the variables are indeed approximately χ^2-distributed), and it is in that case happening only because of random variations. These random variations is what we see in the last point of Fig. 6.17.

8.4. Highest number of clouds

The cases where we find the highest number of clouds giving reliable results were pointed out throughout Section 6.2.

The three figures 6.4, 6.15 and 6.24 are images with constant number of clouds using the highest number clouds for each amplification distribution. The highest number of clouds is 5.4 million for $P_d = 0.0025R_E$ and $P_d = 0.01R_E$, and 9.8 million for $\mu_{max,an} = 50$.

It is interesting to note that we find reliable results for up to 10 million clouds for $\mu_{max,an} = 50$. However, in light of the discussion in Section 8.1, one should be cautious about interpreting this as very significant.

8.5. Level as a function of $\mu_{max,an}$

We examined a situation where the maximum amplification $\mu_{max,an}$ was varied in Section 7.1. $\mu_{max,an}$ was varied from 10 to 120. We see that we have a fairly good fit in Fig. 7.1, and I have created similar plots for this fit as I did in Section 8.2 and included a Q-Q plot for the present case in Fig. 8.3. We see that while there are rather few points, they lie considerably closer to a straight line than in Fig. 8.1.

It is thus with greater confidence we may conclude that

$$T = 2967.8\mu_{max,an}^{0.0176} \tag{8.4}$$

There is still reason to be cautious about claiming that the causal relationship between T and $\mu_{max,an}$ is indeed a power-law, but this may serve as a starting point for further investigation and possible analytical approaches to this problem.

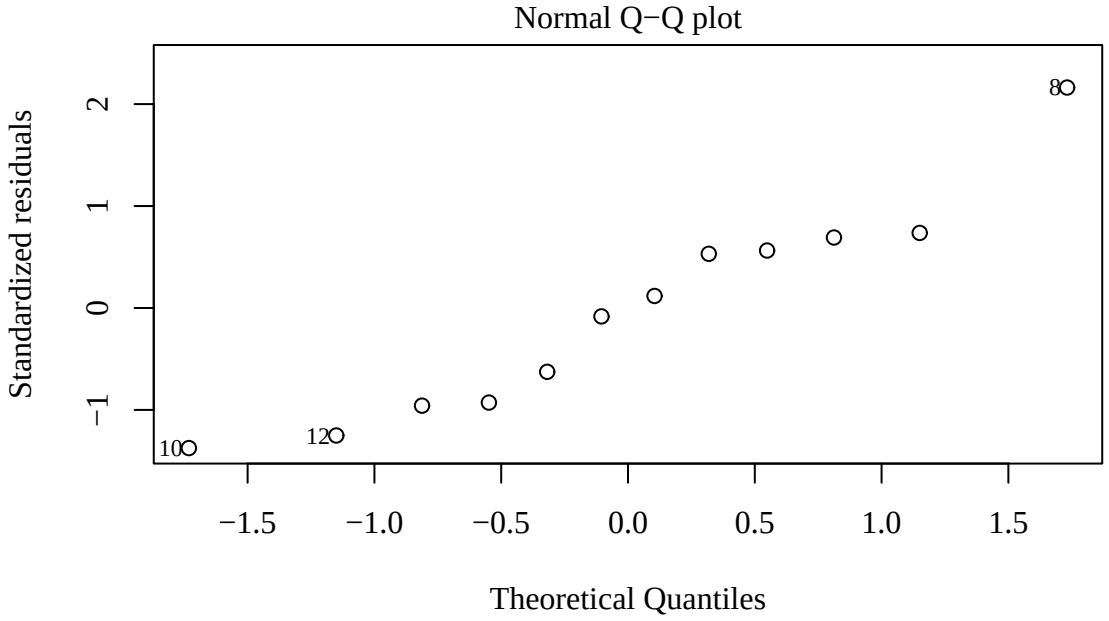

Figure 8.3.: *Q-Q plot for the residuals of the fit of Fig. 7.1. If the errors are normal as assumed, the points would lie on a straight line. The numbers in the plot is the number of the parameter set that the "outlier" stems from.*

8.6. Varying N_{max}

In Section 7.2, we saw how the rejection threshold varies when the maximum number of counts is allowed to increase. I made a linear fit in Fig. 7.2 but, visually, it may seem that the relationship is of a more exponential nature. Indeed, a log-linear fit yielded slightly higher R^2_{adj}. However, the improvement is only so slight, it would be hard to argue that it is significant.

As even fewer points are used in this case than in Section 8.5, it is even harder to discuss the normality assumption. An examination of standard plots revealed nothing that would invalidate this assumption, but as in the previous cases, caution should be exercised when causal relationships are considered.

This relationship can in this case be expressed as

$$T = 0.0025 N_{max} + 300. \tag{8.5}$$

8.7. Two line profiles

We found in Section 7.3 levels for a situation where only two line profiles where used for $N_b = 1000$, $N_c = 1$ million and the $\mu_{max,an} = 50$ amplification distribution, and for both $EW = 8$ nm and $EW = 25$ nm. We found $\alpha = 0.096$ and $\alpha = 1.37 \times 10^{-9}$. This

is substantially higher than $\alpha = 0.012$ and $\alpha = 5.07 \times 10^{-28}$ which is what we find in the case where four line profiles are used. It is only to be expected that it gets harder to detect when fewer line profiles are used.

8.8. *A posteriori* conclusions

I have strongly emphasized that the type of conclusion drawn from this study is an *a priori* conclusion that only make statements about under what conditions lensing may be detected. However, much of the methodology employed in this thesis can be used to make *A posteriori* conclusions as well: When observations are done, what conclusion should be drawn? Let us again turn our attention to Fig. 5.4.

We know that when an observational campaign is completed we will know the values of N_b and EW. Let us for the sake of the argument say that $N_b = 1000$ and $EW = 25$ nm, but the value of N_c is unknown. Also, the value of the test statistic Θ is computed based on the line profiles. We note that the two upper panels of Fig. 5.4 are for this situation, and that the position of the blue histogram depends on N_c.

Now, we may ask the question, when Θ has been measured, can we detect lensing, and if so, what is the most probable value for N_c?

Say that we have measured $\Theta = 4300$. Inspecting the upper panel of Fig. 5.4 and comparing with Table 6.2 we see that this is close to the peak of the blue histogram, which is for $N_c = 1$ million, and well above the rejection threshold at $T = 3613$, and thereby we find that lensing can be said to be detected and that 1 million clouds is the most probable value for N_c. For lower values of Θ, we find that higher values of N_c are the most probably, until about $\Theta = 3100$, where we can no longer claim to detect lensing.

However, N_c is not the only parameter that needs to be constrained: As we have seen, the choice of amplification distribution (we used $P_d = 0.0025 R_E$ in this example) is very important. Similar procedures must be employed to constrain these parameters as well.

Finally, we must consider the possibility that Θ is out of the range of the histograms. For example, what will the conclusion be if $\Theta = 10000$? It is more likely that there is lensing than not, but it is a case that cannot be accounted for by any of the simulations, for any parameter set. It is therefore a signal that is stronger than can be explained by this model, and the underlying assumptions must be questioned. We may say that the model has been falsified.

If, on the other hand, $\Theta < 2500$, it is most likely not lensing, but again, the core assumptions must be questioned, and the model is not likely to be correct.

8.9. A fixed threshold

An investigator may find the present study difficult to use for practical purposes since the thresholds presented are strongly variable. That the threshold depends on N_b and EW is not problematic, since these parameters are known when observations are completed. However, given observations, what threshold should one use to determine if lensing can be detected or not? Clearly, this depends on N_c and the choice of amplification distribution, but it may well be that we have no reliable knowledge as to the value or form of these parameters.

One may therefore desire that the optimal threshold is abandoned for a single threshold that is insensitive to these unknown parameters. This is of course possible, but as we noted that it would be problematic to use a conventional 1, 5, or 10 % significance level, using a single threshold would also come at a prize.

We know that the "equal tails" optimalization of the threshold means that the level α (the probability of making a Type I Error) and the power ϕ (the probability of making a Type II Error) are equal. We have further said the test is reliable if they both are below 0.1. Is it possible to find cases where the level can be fixed to a value below 0.1, and yet the power will not exceed 0.1 for any N_c?

Let us return to Fig. 5.4 and for example use the threshold in the middle panel (which is for $N_b = 1000$, $EW = 25$ nm and $N_c = 5$ million). According to Table 6.2, $\alpha = \phi = 0.052$ in this case. As we see that the distribution of Θ is quite insensitive to N_c in the *unlensed* case, the level of the test would become $\alpha \simeq 0.05$ in the case for $N_c = 1$ million (upper panel), the same as in the $N_c = 5$ million case. Also, the power would become even lower. The level has become considerably larger than when we used the optimized threshold, *but* we see that we still have both the level and the threshold below 0.1, so we have indeed found a way to find a level that is not varying with N_c. An investigator may thus decide that the prize of a worsened level is worth paying, depending on the application.

Obviously, this procedure is only interesting to apply when the level is at most 0.1, that is, under the conditions where the core analysis has found the test to be reliable.

We used the level for $N_c = 5$ million in the above example, but that was because it was the highest N_c we had reliable results for in this example. It is clear that we may want to use a $\alpha = 0.1$ level and find the highest number of clouds that this approach may be used for. As a final note, we may in fact utilize the linear regression of Section 6.3 for this purpose.

8.10. Weaknesses of this investigation

8.10.1. Lines with different continuum levels

The most apparent weakness of this study is that it is not immediately applicable to observations. The main reason for this is that all line profiles are as a first order approximation equal. In other words, the brightness of the quasar continuum is the same for all the images.

It is easy to see why it would have been desirable to allow different continuum levels: Most multiply imaged quasars have different brightness for the images due to different magnification from both macrolensing and microlensing. Applying this investigation directly to an object then relies on a fortunate case where the total amplification adds up to be equal for all the images.

For this reason, I have written an `R` function that may be used to generate line profiles with different continuum levels. This function is named **lines**.**profile** and is included in Appendix A.2. Note that it is not intended as a rigorous implementation of the problem, and that it has not undergone extensive validation. Also, it is perhaps not as accessible as the other code I have written.

I have also reviewed the literature, and found that there exists methods used for purposes similar to those discussed in this section, notably that which seems first used by Atwood et al. (1982). I have, however, not seen that this method has been subject to extensive evaluation, which is in itself a formidable task that I have not sought to undertake. Thus, I do not use this method.

The data is of a more diverse nature in this case than in the case where the line profiles are similar. This is not necessarily negative, as it is often possible to gain more information from statistical analysis in such cases.

However, I am quite content with the focus I have had for this thesis, namely a study of a worst-case scenario, so I have defined this possible future course as out of the scope of the present work.

8.10.2. The poor correlations

Except for the case where a point source is used, I must admit that the arguments of Section 5.3 has a certain scent of “hand-waving”.

However, it is important to emphasize that it is only the most extreme cases we should hope to identify this way, and that has been accomplished.

Also, the investigation does not depend critically on this particular analysis, as the main assessment of the test’s applicability consists of comparing known unlensed cases with known lensed cases.

8.10.3. Highly variable thresholds

As pointed out many times, the thresholds presented are highly variable, and depend strongly on parameters that we have no control over, and which is hard to measure at all. Thus, knowing what threshold to use is difficult.

This is not a weakness of the study *per se*, but rather a consequence of the analysis. I have addressed the problem in Section 8.9, but I emphasize that the primary conclusions are recommendations to observers, which do not depend on a fixed threshold.

8.10.4. Unrealistic physical assumptions

One may argue that the assumptions of Section 4.1 are too unrealistic for the study to be applicable in a real-world situation. To that, I may respond that while they may be unrealistic, it does represent a worst-case scenario, when compared to the previous studies. As previous studies concluded that in the case of random cloud motion, it was unlikely that a significant signal could be discerned, it is this case where I find that for some parameter sets, we may find reliable results.

8.10.5. Bin Width as free parameter

I have realized that it might have been better to use the bin width W_b as free parameter rather than the number of bins. While it would make no difference in the final conclusions. Since the lines scale after the number of bins, using the bin width might have been a stronger aid to intuition.

8.10.6. Too few raymaps?

As noted, the difference between $\mu_{max,an} = 50$ and $P_d = 0.0025R_E$ is greater than between $P_d = 0.0025R_E$ and $P_d = 0.01R_E$. This is somewhat unexpected, and prompted by the fact that the mean of the raymaps for $P_d = 0.0025R_E$ was slightly lower than desired. I speculated if the reason might be that I had too few raymaps in spite of the reasonably good convergence of the mean. It is not unthinkable that there may be too few cases with very high amplification, so that the highest amplification was not well represented by the Probability Function. I therefore made some preliminary simulations where I doubled the number of raymaps, found that the mean improved and there was a reasonably high number of cases with very high amplification. However, this had no appreciable impact on the rejection threshold, and therefore I did not do a full run with the new Probability Function, as I am confident that the number of raymaps in the first case is sufficient.

8.11. Summary of conclusions

- If one can assume that the quasar has approximately 1 million clouds, there is good chance to detect lensing, even with a modest line and a rather few bins.
- If one can assume a larger number of clouds, one must choose the line and the instrument with more care.
- If one has no knowledge as to the number of clouds, one should choose instruments and objects that can provide the strongest lines and resolve it into as many as $N_b = 1000$ bins.
- Linear regression, which should be used with great caution due to non-normality of residuals yield the following relations for the threshold T in the core set, for $P_d = 0.0025R_E$, $P_d = 0.01R_E$ and $\mu_{max,an} = 50$ respectively:
  - $T = 4.84EW + 0.127N_b - 71.3\log(N_c) + \mathrm{df} + 1015$,
  - $T = 4.32EW + 0.114N_b - 63.5\log(N_c) + \mathrm{df} + 904$ and
  - $T = 7.57EW + 0.199N_b - 111.4\log(N_c) + \mathrm{df} + 1585$.
- The threshold can also be found by linear regression to follow the maximum amplification as
$$T = 2967.8\mu_{max,an}^{0.0176}.$$
- Further, the threshold can be found by linear regression to follow the maximum number of photons in a bin as
$$T = 0.0025N_{max} + 300.$$
In the last two cases, the intercept is strongly connected to the degrees of freedom in the test used in the example.
- It is significantly more difficult to detect lensing for objects with two quasar images than four.
- It is possible to use these results to draw conclusions after observations have been made, in addition to the recommendations made above.

I have been modest and cautious in this discussion, but I shall now end by again pointing out that it is a *worst-case* scenario that has been studied in this thesis, and so, under the conditions where I have concluded that significant results may be obtained, it is quite likely that an empirical investigation will yield good results.

Appendix

A. R-code

Note that the variable (object) names often bear testimony of the author's changing thoughts and visions during this project, and do not as accurately represent the concepts as the text of this thesis.

A.1. Amplification

```
pamp <- function(mu, data, ...) UseMethod("pamp", data)
damp <- function(mu, data, ...) UseMethod("damp", data)
qamp <- function(p, data, ...) UseMethod("qamp", data)
ramp <- function(n, data, ...) UseMethod("ramp", data)

# Distribution functions for amplification
pamp.default <- function(mu, data)
  ifelse(mu >= 1, 1, 0)

pamp.point <- function(mu, data)
  ifelse(mu >= 0.5,
         1 - 0.25 * mu^-2,
         0)

pamp.cutoff <- function(mu, data)
{
  if (data$mulim > 1)
  {
    mumax <- data$mulim
    mumin <- (data$mulim / (2 * data$mulim - 1))
  }
  else if (data$mulim > 0.50005 & data$mulim < 1)
  {
    mumin <- data$mulim
    mumax <- (data$mulim / (2 * data$mulim - 1))
  }
  else
```

```
  {
    stop("mulim <= 0.50005")
  }
  return(ifelse(mu < mumin,
                0,
                ifelse(mu >= mumax,
                       1,
                       1 / (4 * mumin * (1 - mumin)) * (1 - mumin^2 / mu^2))))
}

# Probability density functions for amplification
damp.default <- function(mu, data)
  ifelse(mu == 1, 1, 0)

damp.point <- function(mu, data)
  ifelse(mu >= 0.5,
         0.5 * mu^-3,
         0)

damp.cutoff <- function(mu, data)
{
  if (data$mulim > 1)
  {
    mumax <- data$mulim
    mumin <- (data$mulim / (2 * data$mulim - 1))
  }
  else if (data$mulim > 0.50005 & data$mulim < 1)
  {
    mumin <- data$mulim
    mumax <- (data$mulim / (2 * data$mulim - 1))
  }
  else
  {
    stop("mulim <= 0.50005")
  }
  return(ifelse(mu < mumin | mu >= mumax,
                0,
                mumin / (2 * (1 - mumin) * mu^3)))
}

# Quantile functions for amplification
qamp.default <- function(p, data) 1

qamp.point <- function(p, data)
```

```
  # Note that a loss-of-precision problem may occur here
   (1 / (2 * sqrt(1 - p)))

qamp.cutoff <- function(p, data)
{
  mumin <- if(data$mulim > 1)
               (data$mulim / (2 * data$mulim - 1))
           else if (data$mulim > 0.50005 & data$mulim < 1)
                    data$mulim
           else stop("mulim <= 0.50005")
  mumin / sqrt(1 - (4 * p * mumin * (1 - mumin)))
}


# Generates random variates for amplification
ramp.default <- function(n, data) rep(1, n)

ramp.point <- function(n, data)
{
  u <- runif(n)
  # Because of the finite representation of numbers,
  # the interval must be split.
   splitpt <- 0.9999999 # This is splitting on the ten millionth quantile.
  new <- which(u > splitpt)
  u[new] <- runif(length(new), splitpt, 1)
  return(qamp(u, data))
}

ramp.cutoff <- function(n, data)
{
  # sending the right data object will return the right quantile function
  return(ramp.point(n, data))
}

ramp.histogram <- function(n, data)
  sample(data$mids, n, replace=T, data$counts/sum(data$counts))
```

A.2. Line Generation

```
ncloudsbin <- function(binno, ntotalclouds, numberofbins, linewidth)
# using binomial would be rbinom(1, ntotalclouds, dnorm(binno, sd = linewidth))
  return(rpois(1, ntotalclouds * dnorm(binno, sd = linewidth)))

fluxfromcloud <- function(ncloudsbin, cloudintensity, ampdata, exptime)
```

```
  return(if(ncloudsbin == 0) 0
  else cloudintensity * exptime * ramp(ncloudsbin, ampdata))


photonsdetectedbin <- function(fluxfromcloud, ntotalclouds,
                               continuumintensity, sensitivity,
                               exptime)
  # Bug here in R version < 1.2.0
  return(sum(rpois(length(fluxfromcloud), sensitivity
                   * fluxfromcloud))
         # Might use rgeom here
         + rpois(1, sensitivity * exptime * continuumintensity))

lineprofiles <- function(ntotalclouds, EW, numberofbins, ampdata,
                         fluxdensity=3e-18, maxcount=20000, binwidth=0.1,
                         exptime=NULL, aperture=8, sensitivity=0.1,
                         nprofiles=4, ...)
{
   nullprofile <- sapply(-as.integer(numberofbins/2):as.integer(numberofbins/2),
                         ncloudsbin, ntotalclouds, numberofbins,
                         numberofbins/7)
  continuumintensity <- nophot(fluxdensity, 1, binwidth,
                               pi * (aperture/2)^2, ...)
   cloudintensity <- continuumintensity * EW /( binwidth *
                         ntotalclouds)
  if(is.null(exptime)) exptime <- maxcount / ( sensitivity *
                           ( max(nullprofile) *
                            cloudintensity + continuumintensity))
  apply(matrix(nullprofile, nprofiles, length( nullprofile ), byrow=T),
                c(1,2), function(x) photonsdetectedbin(
                      fluxfromcloud(x, cloudintensity, ampdata, exptime),
                      ntotalclouds, continuumintensity,
                         sensitivity, exptime))
}


lines.profile <- function(ntotalclouds, EW, numberofbins, ampdata,
                         fluxdensity=3e-18, maxcount=20000, binwidth=0.1,
                         exptime=NULL, aperture=8, sensitivity=0.1,
                         nprofiles=4, ...)
{
   nullprofile <- sapply(-as.integer(numberofbins/2):as.integer(numberofbins/2),
                         ncloudsbin, ntotalclouds, numberofbins,
```

```
                                numberofbins/7)
  if (length(fluxdensity) <= 1) fluxdensity <- fluxdensity *
                                (max(EW)/EW)
  continuumintensity <- sapply(fluxdensity, nophot, 1, binwidth,
                                     pi * (aperture /2)^2, ...)
  cloudintensity <- continuumintensity * EW /( binwidth *
                                ntotalclouds)
  if (is.null(exptime)) exptime <- maxcount / ( sensitivity *
                                  ( max(nullprofile ) *
                                   max(cloudintensity) +
                                   max(continuumintensity)))
  matr <- cbind(rep(continuumintensity, length=nprofiles),
                rep(cloudintensity , length=nprofiles),
                matrix( nullprofile , nprofiles ,
                                    length( nullprofile ), byrow=T))
  t(apply(matr, 1,
        function(row) sapply(row[3:length(row)], function(x) {
                 photonsdetectedbin(
                        fluxfromcloud(x , row[2], ampdata, exptime),
                        ntotalclouds , row[1], sensitivity , exptime)})))
}


# Ad hoc function for identifying bins with amplification
fluxfromcloudwa <- function(ncloudsbin, cloudintensity, ampdata, exptime)
{
  if (ncloudsbin == 0) {
    flux <- 0
    write(c(NA, NA, NA), "~/hfag/data/amps1M251000ex200t25.dat", append=T)
  } else {
    amps <- ramp(ncloudsbin, ampdata)
    write(c(max(amps), max(abs(log10(amps))), mean(amps - 1)),
          "~/hfag/data/amps1M251000ex200t25.dat", append=T)
    flux <- cloudintensity * exptime * amps
  }
return(flux)
}
```

A.3. Interface to `cmap`

```
readmap <- function(file, normalized=TRUE, asvector=FALSE, ...)
{
  i <- as.numeric(sub(".*i_([0-9]*)_.*", "\\1", file ))
  j <- as.numeric(sub(".*j_([0-9]*)_.*", "\\1", file ))
```

```
  map <- readBin(file, "double", n=i*j , size=4)
  if (normalized)
  {
    Q <- as.numeric(sub(".*Q_([0-9\.]*)_.*", "\\1", file))
    map <- map / Q
  }
  if (! asvector) map <- matrix(map, i, j)
  return(map)
}

# The below function runs Stein Vidar's cmap.
runcmap <- function(Kappa=0.29, Gamma=0, Theta=0, Quality=10, i=256, j=256, xTarget=10, yTarget=10
  system(paste("cmap -K", Kappa, "-G", Gamma, "-T", Theta, "-Q", Quality, "-i", i, "-j", j, "-X", xTarget

summarymap <- function(file, relhistresolution=0.1, normalized=F, lg="",
                       ...)
{
  map <- readmap(file, normalized, asvector=TRUE, ...)
  mmean <- mean(map)
  mmax <- max(map)
  mmin <- min(map)
  mvar <- var(map)
  Q <- as.numeric(sub(".*Q_([0-9\.]*)_.*", "\\1", file))
  seed <- as.integer(sub(".*s_([0-9]*)\..*", "\\1", file ))
  res <- (if (normalized) 1/(Q * relhistresolution )
          else (1 / relhistresolution ))
#browser()
  breaks <- if(lg != "xy")
    seq(0, round.multiple(mmax, res, up=T), res)
  else
    seq(-2, ceiling(log(mmax)), 10 * res)
  mhist <- hist(map, breaks=breaks, plot=F, log=lg, ...)
  return( list (histogram=mhist,
                mapmin=mmin,
                mapmax=mmax,
                mapvar=mvar,
                mapmean=mmean,
                mapseed=seed))
}

vecidmaps <- function(path, files, quiet=F , ...)
{
  filearr <- list. files (path, files , full .names=T)
  if (! quiet)
```

```
  {
    print(noquote("Files found:"))
    print(noquote(filearr))
  }
  return(apply(cbind(filearr), 1, readmap, ...))
}

# The below is a function used only in a debugging session
findsmall <- function(path, files, smallest=0.25, quiet=F, ...)
{
  vec <- sapply(list.files(path, files, full.names=T), function(file)
                  if (any(readmap(file, normalized=T, asvector=F) < smallest))
                  file else NA)
  vec[!is.na(vec)]
}

summaryidmaps <- function(path, files, quiet=F, lg="", ...)
{
   filearr <- list.files(path, files, full.names=T)
  tmp <- summarymap(filearr[1], lg=lg, ...)
  if (! quiet) print( filearr [1])
  h <- tmp$histogram
  sm <- tmp$mapmean
  sv <- tmp$mapvar
  mmin <- tmp$mapmin
  mmax <- tmp$mapmax
  seed <- tmp$mapseed
  nfiles <- length(filearr)
  for ( i in 2: nfiles ) # Supposedly, reduced efficiency by using a for
  {                      # loop is neglible compared to file access time.
    tmp <- summarymap(filearr[i], lg=lg , ...)
    if (! quiet) print( filearr [ i ])
    h <- add.histogram(h, tmp$histogram)
    sm <- c(sm, tmp$mapmean)
    sv <- c(sv, tmp$mapvar)
    seed <- c(seed, tmp$mapseed)
    mmin <- min(c(mmin, tmp$mapmin))
    mmax <- max(c(mmax, tmp$mapmax))
  }
  if (! quiet) print(noquote(paste(nfiles, " files read." )))
  return( list (histogram=h,
                nfiles = nfiles ,
                mapmin=mmin,
                mapmax=mmax,
```

```
                mapvars=sv,
                mapmeans=sm,
                mapseed=seed))
}


# Function takes a _normalized_ raymap as input and computes a
# lightcurve with a circular source with a radius of *radius* _pixels_
# (meaning, you need to know the pixel size in Einstein radii),
# centred at a row or col.
lightcurve <- function(map, radius=1, row=dim(map)[1]%/%2, col=NULL)
{
  if (mean(map) > 1.1)
    warning(paste("Is", deparse(substitute(map)), "really normalized?"))
  map <- t(map)
  if (!is.null(col)) row <- col
  if (row > dim(map)[1] - radius || row < radius)
    stop("row/col is outside the raymap!")
  radius <- radius - 1
  means <- sapply((radius + 1):(dim(map)[2] - radius), function(x) {
    indices <- sapply(round(x - radius):round(x + radius), function(i) {
      tmp <- round(sqrt(radius^2 - abs(x - i)^2))
      cbind((-tmp:tmp) + row, i) }, simplify=F)
    mean(unlist(sapply(indices, function(i) map[i[1], i[2]])) })
  c(rep(NA, radius), means, rep(NA, radius))

}
```

A.4. Miscellaneous

```
mag2flux <- function(mag, fluxvega=3.5e-11)
  return(fluxvega * 10^(- mag / 2.5))

flux2mag <- function(flux, fluxvega=3.5e-11)
  return(-2.5 * log10(flux/fluxvega))


nophot <- function(flux, exptime, binwidth, aperturearea, wavelength=550e-9,
                   lightspeed=299792458, planck=6.626e-34)
  return(flux * wavelength * exptime * binwidth * aperturearea /
         (planck * lightspeed))
```

B. Extended tables of results

This appendix has tables of the parameter sets. Note that all tables extend over several pages.

Also note that I use a reliability level $\alpha_r = 0.125$ in this tables as opposed to $\alpha_r = 0.1$ for all other applications in this thesis. The reason for this is that one may be interested to study parameter sets that is slightly above the before mentioned level.

Table B.1.: *Rejection thresholds (T), levels (α), mean χ^2-test statistics ($\bar{\Theta}$) and p-values for the $P_d = 0.01R_E$ amplification distribution. The number of tests conducted for that parameter set is also tabulated. The reliability level is $\alpha_r = 0.125$.*

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
19	1000000	18.2	100	366	0.043	329	0.117	213
25	1000000	21.6	100	373	0.028	331	0.106	213
31	1000000	25.0	100	373	0.034	331	0.103	713
37	1000000	8.0	250	842	0.049	798	0.109	188
43	1000000	11.4	250	898	0.008	808	0.070	188
49	1000000	14.8	250	940	0.001	848	0.007	188
55	1000000	18.2	250	971	0.000	847	0.008	188
61	1000000	21.6	250	1014	0.000	881	0.001	188
67	1000000	25.0	250	1025	0.000	844	0.009	188
68	3200000	25.0	250	841	0.046	795	0.124	188
73	1000000	8.0	400	1339	0.035	1264	0.096	162
79	1000000	11.4	400	1422	0.001	1310	0.014	162
85	1000000	14.8	400	1493	0.000	1320	0.008	162
91	1000000	18.2	400	1568	0.000	1363	0.001	162
92	3200000	18.2	400	1313	0.057	1259	0.114	162
97	1000000	21.6	400	1647	0.000	1385	0.000	162
98	3200000	21.6	400	1328	0.040	1263	0.100	162
103	1000000	25.0	400	1691	0.000	1397	0.000	162

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
104	3200000	25.0	400	1361	0.015	1278	0.057	162
109	1000000	8.0	550	1820	0.019	1733	0.075	150
115	1000000	11.4	550	1921	0.001	1747	0.048	150
121	1000000	14.8	550	2037	0.000	1815	0.003	150
127	1000000	18.2	550	2147	0.000	1872	0.000	150
128	3200000	18.2	550	1803	0.040	1723	0.103	150
133	1000000	21.6	550	2248	0.000	1919	0.000	150
134	3200000	21.6	550	1849	0.013	1721	0.108	150
139	1000000	25.0	550	2342	0.000	1968	0.000	150
140	3200000	25.0	550	1866	0.007	1758	0.031	150
145	1000000	8.0	700	2286	0.030	2179	0.112	150
151	1000000	11.4	700	2413	0.001	2253	0.010	150
157	1000000	14.8	700	2566	0.000	2324	0.000	150
163	1000000	18.2	700	2703	0.000	2393	0.000	150
164	3200000	18.2	700	2291	0.022	2197	0.068	150
169	1000000	21.6	700	2827	0.000	2436	0.000	150
170	3200000	21.6	700	2332	0.010	2206	0.052	150
175	1000000	25.0	700	2974	0.000	2466	0.000	150
176	3200000	25.0	700	2373	0.002	2243	0.015	150
181	1000000	8.0	850	2742	0.029	2651	0.079	150
187	1000000	11.4	850	2897	0.001	2727	0.008	150
193	1000000	14.8	850	3054	0.000	2819	0.000	150
199	1000000	18.2	850	3214	0.000	2835	0.000	150
200	3200000	18.2	850	2749	0.034	2648	0.087	150
205	1000000	21.6	850	3399	0.000	2981	0.000	150
206	3200000	21.6	850	2818	0.007	2672	0.045	150
211	1000000	25.0	850	3578	0.000	2986	0.000	206
212	3200000	25.0	850	2874	0.002	2702	0.018	206
213	5400000	25.0	850	2743	0.039	2647	0.088	206
217	1000000	8.0	1000	3202	0.038	3095	0.110	580
223	1000000	11.4	1000	3380	0.000	3203	0.005	150
229	1000000	14.8	1000	3562	0.000	3312	0.000	150
230	3200000	14.8	1000	3180	0.048	3090	0.122	150
235	1000000	18.2	1000	3748	0.000	3324	0.000	150
236	3200000	18.2	1000	3247	0.016	3119	0.064	150
241	1000000	21.6	1000	3963	0.000	3468	0.000	150
242	3200000	21.6	1000	3306	0.004	3153	0.025	150

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
243	5400000	21.6	1000	3178	0.055	3099	0.101	150
247	1000000	25.0	1000	4169	0.000	3508	0.000	680
248	3200000	25.0	1000	3360	0.001	3159	0.022	180
249	5400000	25.0	1000	3220	0.036	3100	0.099	180

Table B.2.: *Rejection thresholds (T), levels (α), mean χ^2-test statistics ($\bar{\Theta}$) and p-values for the $P_d = 0.0025R_E$ amplification distribution. The number of tests conducted for that parameter set is also tabulated. The reliability level is $\alpha_r = 0.125$.*

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
13	1000000	14.8	100	364	0.043	330	0.114	213
19	1000000	18.2	100	371	0.036	333	0.090	213
25	1000000	21.6	100	380	0.016	335	0.078	213
31	1000000	25.0	100	384	0.015	339	0.062	713
37	1000000	8.0	250	867	0.024	808	0.071	188
43	1000000	11.4	250	921	0.003	827	0.026	188
49	1000000	14.8	250	956	0.001	850	0.006	188
55	1000000	18.2	250	998	0.000	862	0.003	188
61	1000000	21.6	250	1038	0.000	884	0.001	188
62	3200000	21.6	250	841	0.058	795	0.123	188
67	1000000	25.0	250	1074	0.000	883	0.001	188
73	1000000	8.0	400	1350	0.019	1274	0.068	162
79	1000000	11.4	400	1435	0.002	1299	0.023	162
85	1000000	14.8	400	1529	0.000	1334	0.004	162
91	1000000	18.2	400	1624	0.000	1366	0.001	162
92	3200000	18.2	400	1331	0.034	1268	0.086	162
97	1000000	21.6	400	1689	0.000	1400	0.000	162
98	3200000	21.6	400	1354	0.016	1271	0.075	162
103	1000000	25.0	400	1767	0.000	1450	0.000	162
104	3200000	25.0	400	1377	0.011	1274	0.067	162
109	1000000	8.0	550	1840	0.018	1741	0.058	150
115	1000000	11.4	550	1954	0.000	1767	0.023	150
121	1000000	14.8	550	2071	0.000	1797	0.006	150
127	1000000	18.2	550	2210	0.000	1892	0.000	150
128	3200000	18.2	550	1825	0.012	1747	0.048	150
133	1000000	21.6	550	2318	0.000	1979	0.000	150
134	3200000	21.6	550	1876	0.006	1729	0.087	150
139	1000000	25.0	550	2441	0.000	2019	0.000	150
140	3200000	25.0	550	1894	0.002	1768	0.022	150
141	5400000	25.0	550	1795	0.044	1721	0.110	150
145	1000000	8.0	700	2311	0.010	2203	0.057	150
151	1000000	11.4	700	2454	0.000	2263	0.007	150
157	1000000	14.8	700	2610	0.000	2321	0.000	150

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
163	1000000	18.2	700	2770	0.000	2401	0.000	150
164	3200000	18.2	700	2323	0.008	2213	0.042	150
169	1000000	21.6	700	2935	0.000	2466	0.000	150
170	3200000	21.6	700	2366	0.004	2220	0.034	150
175	1000000	25.0	700	3090	0.000	2551	0.000	150
176	3200000	25.0	700	2406	0.001	2254	0.010	150
177	5400000	25.0	700	2289	0.024	2182	0.104	150
181	1000000	8.0	850	2771	0.019	2668	0.051	150
187	1000000	11.4	850	2940	0.000	2766	0.002	150
193	1000000	14.8	850	3120	0.000	2828	0.000	150
194	3200000	14.8	850	2726	0.053	2638	0.110	150
199	1000000	18.2	850	3315	0.000	2931	0.000	150
200	3200000	18.2	850	2795	0.010	2680	0.036	150
205	1000000	21.6	850	3507	0.000	3004	0.000	150
206	3200000	21.6	850	2838	0.008	2675	0.041	150
211	1000000	25.0	850	3698	0.000	3038	0.000	206
212	3200000	25.0	850	2910	0.000	2745	0.004	206
213	5400000	25.0	850	2776	0.016	2670	0.048	206
217	1000000	8.0	1000	3233	0.019	3118	0.065	580
223	1000000	11.4	1000	3417	0.000	3211	0.004	150
229	1000000	14.8	1000	3623	0.000	3304	0.000	150
230	3200000	14.8	1000	3193	0.045	3092	0.119	150
235	1000000	18.2	1000	3852	0.000	3397	0.000	150
236	3200000	18.2	1000	3273	0.013	3148	0.030	150
241	1000000	21.6	1000	4083	0.000	3529	0.000	150
242	3200000	21.6	1000	3344	0.001	3169	0.016	150
243	5400000	21.6	1000	3205	0.036	3104	0.092	150
247	1000000	25.0	1000	4307	0.000	3542	0.000	680
248	3200000	25.0	1000	3406	0.000	3166	0.018	180
249	5400000	25.0	1000	3246	0.015	3126	0.053	180

Table B.3.: *Rejection thresholds (T), levels (α), mean χ^2-test statistics ($\bar{\Theta}$) and p-values for the $\mu_{max,an} = 50$ amplification distribution. The number of tests conducted for that parameter set is also tabulated. The reliability level is $\alpha_r = 0.125$.*

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
1	1000000	8.0	100	367	0.040	332	0.099	713
7	1000000	11.4	100	385	0.012	339	0.060	213
13	1000000	14.8	100	398	0.006	348	0.030	213
19	1000000	18.2	100	412	0.003	356	0.015	213
25	1000000	21.6	100	421	0.001	359	0.011	213
31	1000000	25.0	100	430	0.001	355	0.016	713
37	1000000	8.0	250	923	0.003	827	0.026	188
43	1000000	11.4	250	1000	0.000	837	0.014	188
49	1000000	14.8	250	1076	0.000	893	0.000	188
50	3200000	14.8	250	855	0.046	799	0.104	188
55	1000000	18.2	250	1133	0.000	919	0.000	188
56	3200000	18.2	250	870	0.023	801	0.097	188
61	1000000	21.6	250	1195	0.000	942	0.000	188
62	3200000	21.6	250	893	0.005	820	0.039	188
67	1000000	25.0	250	1240	0.000	961	0.000	188
68	3200000	25.0	250	903	0.007	816	0.046	188
73	1000000	8.0	400	1441	0.001	1315	0.011	162
79	1000000	11.4	400	1575	0.000	1378	0.000	162
80	3200000	11.4	400	1320	0.047	1265	0.092	162
85	1000000	14.8	400	1714	0.000	1420	0.000	162
86	3200000	14.8	400	1364	0.009	1277	0.060	162
91	1000000	18.2	400	1830	0.000	1485	0.000	162
92	3200000	18.2	400	1402	0.004	1295	0.029	162
93	5400000	18.2	400	1326	0.045	1263	0.099	162
97	1000000	21.6	400	1957	0.000	1482	0.000	162
98	3200000	21.6	400	1438	0.001	1301	0.022	162
99	5400000	21.6	400	1341	0.027	1264	0.097	162
103	1000000	25.0	400	2052	0.000	1573	0.000	162
104	3200000	25.0	400	1470	0.000	1324	0.007	162
105	5400000	25.0	400	1355	0.018	1273	0.070	162
107	9800000	25.0	400	1293	0.083	1257	0.122	162
109	1000000	8.0	550	1936	0.001	1762	0.028	150
115	1000000	11.4	550	2133	0.000	1857	0.000	150

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
116	3200000	11.4	550	1799	0.040	1722	0.106	150
121	1000000	14.8	550	2315	0.000	1959	0.000	150
122	3200000	14.8	550	1865	0.006	1756	0.034	150
127	1000000	18.2	550	2510	0.000	2021	0.000	150
128	3200000	18.2	550	1916	0.001	1793	0.007	150
129	5400000	18.2	550	1809	0.031	1725	0.098	150
133	1000000	21.6	550	2666	0.000	2118	0.000	150
134	3200000	21.6	550	1963	0.000	1771	0.019	150
135	5400000	21.6	550	1839	0.016	1739	0.062	150
136	7600000	21.6	550	1791	0.049	1723	0.104	150
139	1000000	25.0	550	2859	0.000	2170	0.000	150
140	3200000	25.0	550	2031	0.000	1844	0.001	150
141	5400000	25.0	550	1875	0.004	1760	0.030	150
145	1000000	8.0	700	2409	0.001	2243	0.015	150
151	1000000	11.4	700	2644	0.000	2343	0.000	150
157	1000000	14.8	700	2876	0.000	2462	0.000	150
158	3200000	14.8	700	2344	0.008	2223	0.031	150
159	5400000	14.8	700	2256	0.041	2178	0.116	150
163	1000000	18.2	700	3132	0.000	2590	0.000	150
164	3200000	18.2	700	2429	0.000	2248	0.012	150
165	5400000	18.2	700	2294	0.022	2208	0.049	150
169	1000000	21.6	700	3362	0.000	2672	0.000	150
170	3200000	21.6	700	2499	0.000	2305	0.001	150
171	5400000	21.6	700	2327	0.007	2199	0.065	150
172	7600000	21.6	700	2267	0.040	2178	0.114	150
175	1000000	25.0	700	3603	0.000	2803	0.000	150
176	3200000	25.0	700	2570	0.000	2317	0.001	150
177	5400000	25.0	700	2374	0.002	2210	0.046	150
178	7600000	25.0	700	2300	0.020	2211	0.046	150
181	1000000	8.0	850	2900	0.000	2739	0.005	150
187	1000000	11.4	850	3150	0.000	2875	0.000	150
188	3200000	11.4	850	2750	0.030	2648	0.086	150
193	1000000	14.8	850	3418	0.000	2984	0.000	150
194	3200000	14.8	850	2824	0.005	2663	0.058	150
195	5400000	14.8	850	2724	0.056	2636	0.116	150
199	1000000	18.2	850	3730	0.000	3112	0.000	150
200	3200000	18.2	850	2916	0.001	2705	0.016	150

	N_c	EW	N_b	$\bar{\Theta}$	p-value	T	α	Tests
201	5400000	18.2	850	2764	0.018	2647	0.089	150
205	1000000	21.6	850	4032	0.000	3228	0.000	150
206	3200000	21.6	850	3021	0.000	2794	0.000	150
207	5400000	21.6	850	2824	0.005	2679	0.037	150
208	7600000	21.6	850	2744	0.031	2644	0.095	150
211	1000000	25.0	850	4317	0.000	3262	0.000	206
212	3200000	25.0	850	3099	0.000	2823	0.000	206
213	5400000	25.0	850	2886	0.001	2705	0.017	206
214	7600000	25.0	850	2793	0.011	2660	0.064	206
215	9800000	25.0	850	2738	0.040	2644	0.096	206
217	1000000	8.0	1000	3360	0.001	3179	0.011	580
223	1000000	11.4	1000	3642	0.000	3320	0.000	150
224	3200000	11.4	1000	3200	0.034	3096	0.108	150
229	1000000	14.8	1000	3963	0.000	3475	0.000	150
230	3200000	14.8	1000	3310	0.005	3140	0.037	150
231	5400000	14.8	1000	3180	0.049	3092	0.119	150
235	1000000	18.2	1000	4302	0.000	3633	0.000	150
236	3200000	18.2	1000	3419	0.000	3218	0.003	150
237	5400000	18.2	1000	3249	0.016	3124	0.057	150
241	1000000	21.6	1000	4653	0.000	3813	0.000	150
242	3200000	21.6	1000	3524	0.000	3206	0.005	150
243	5400000	21.6	1000	3292	0.009	3141	0.036	150
244	7600000	21.6	1000	3220	0.026	3093	0.116	150
247	1000000	25.0	1000	5013	0.000	3926	0.000	680
248	3200000	25.0	1000	3626	0.000	3279	0.000	180
249	5400000	25.0	1000	3386	0.000	3194	0.007	180
250	7600000	25.0	1000	3268	0.011	3128	0.051	180
251	9800000	25.0	1000	3206	0.030	3109	0.080	180

Bibliography

Arav N., Barlow T. A., Laor A., Sargent W. L. W., Blandford R. D., 1998, MNRAS 297, 990

Atwood B., Baldwin J. A., Carswell R. F., 1982, ApJ 257, 559

Babu G. J., Feigelson E. D., 1996, Astrostatistics – interdisciplinary statistics, London Chapman and Hall

Bhattacharyya G. K., Johnson R. A., 1977, Statistical Concepts And Methods, John Wiley & Sons

Blandford R., Narayan R., 1986, ApJ 310, 568

Bottorff M., Korista K. T., Shlosman I., Blandford R. D., 1997, ApJ 479, 200

Chang K., Refsdal S., 1979, Nature 282, 561

Done C., Krolik J. H., 1996, ApJ 463, 144

Dudewicz E. J., Mishra S. N., 1988, Modern Mathematical Statistics, Wiley

Emmering R. T., Blandford R. D., Shlosman I., 1992, ApJ 385, 460

Grieger B., Kayser R., Refsdal S., 1986, Nature 324, 126

Haugan S. V. H., 1994, Master's thesis, University of Oslo, Institute of Theoretical Astrophysics

Haugan S. V. H., 1996, in IAU Symp. 173: Astrophysical Applications of Gravitational Lensing, Vol. 173, p. 275

Haugan S. V. H., 1999, MNRAS 303, 471

Haugan S. V. H., 2002, Ph.D. thesis, University of Oslo, Institute of Theoretical Astrophysics

Huchra J., Gorenstein M., Kent S., Shapiro I., Smith G., Horine E., Perley R., 1985, AJ 90, 691

Ihaka R., Gentleman R., 1996, Journal of Computational and Graphical Statistics 5(3), 299

Irwin M. J., Hewett P. C., Corrigan R. T., Jedrzejewski R. I., Webster R. L., 1989, AJ 98, 1989

Kayser R., Refsdal S., 1983, A&A 128, 156

Kayser R., Refsdal S., Stabell R., 1986, A&A 166, 36

Krolik J. H., 1999, Active Galactic Nuclei, Princeton Series in Astrophysics, Princeton University Press, Princeton, New Jersey

Lebach D. E., Corey B. E., Shapiro I. I., Ratner M. I., Webber J. C., Rogers A. E. E., Davis J. L., Herring T. A., 1995, Phys. Rev. Lett. 75, 1439

Lewis G. F., Irwin M. J., 1995, MNRAS 276, 103

Lewis G. F., Irwin M. J., Hewett P. C., Foltz C. B., 1998, MNRAS 295, 573

Lewis G. F., Miralda-Escude J., Richardson D. C., Wambsganss J., 1993, MNRAS 261, 647

Álvarez M. C., 2000, Private communication, E-mail correspondence.

Narlikar J. V., 1993, Introduction to Cosmology, Cambridge University Press, 2nd edition

Nemiroff R. J., 1988, ApJ 335, 593

Paczynski B., 1986, ApJ 301, 503

Refsdal S., 1964, MNRAS 128, 307

Refsdal S., 1970, ApJ 159, 357

Refsdal S., Surdej J., 1994, Reports on Progress in Physics 57(2), 117

Robson I., 1996, Active galactic nuclei, Wiley-Praxis Series in Astronomy and Astrophysics, New York, NY: Wiley, Chichester: Praxis Publishing

Schneider P., Ehlers J. ., Falco E. E., 1992, Gravitational Lenses, Springer-Verlag Berlin Heidelberg New York. Also Astronomy and Astrophysics Library

Schneider P., Wambsganss J., 1990, A&A 237, 42

Schneider P., Weiss A., 1987, A&A 171, 49

Sulentic J. W., Marziani P., Dultzin-Hacyan D., 2000, ARA&A 38, 521

Walsh D., Carswell R. F., Weymann R. J., 1979, Nature 279, 381

Wambsganss J., Schneider P., Paczynski B., 1990, ApJ 358, L33

Witt H. J., 1993, ApJ 403, 530

Index